25/1.9.

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Volume 59, 1960

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

Second-class postage paid at Lancaster, Pa.

© 1960 by the American Psychological Association, Inc.

Bureau Edn. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated 1.9.70
Acc. No. J 253

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 59

Adams, P. A. See Postman, L.
Adamson, R. See Bevan, W.
Anderson, N. H. Effect of First-Order Conditional Probability in a Two-Choice Learning Situation 73
Archer, E. J. See Namikas, G.
Atthowe, J. M., Jr. Types of Conflict and Their Resolution 1
Bardach, J. L. Effects of Situational Anxiety at Different Stages of Practice 420
Beeman, E. Y., Hartman, T. F., and Grant, D. A. Supplementary Report: Influence of Intertrial Interval during Extinction on Spontaneous Recovery of Conditioned Eyelid Responses 279
Bergum, B. O. Gradients of Generalization in Secondary Reinforcement 47
Berry, R. N., and Davis, R. C. The Somatic Background of Rote Learning 27
Bevan, W., and Adamson, R. Reinforcers and Reinforcement: Their Relation of Maze Performance 226
Bilodeau, E. A., and Ryan, F. J. A Test for Interaction of Delay of Knowledge of Results and Two Types of Interpolated Activity 414
Botwinick, J., Robbin, J. S., and Brinley, J. F. Age Differences in Card-Sorting Performance in Relation to Task Difficulty, Task Set, and Practice 10
Bower, G. H. Partial and Correlated Reward in Escape Learning 126
Brinley, J. F. See Botwinick, J.
Brogden, W. F. See Ernst, R. L.
Brogden, W. J. See Hoffeld, D. R.
Brogden, W. J. See Namikas, G.
Brown, J. Evidence for a Selective Process during Perception of Tachistoscopically Presented Stimuli 176
Cahill, H. E., and Hovland, C. I. The Role of Memory in the Acquisition of Concepts 137
Camp, B. W. "Association" of Nonsense Syllables following Varied Learning Conditions 35
Champion, R. A. See Standish, R. R.
Coleman, E. B. See DeSoto, C. B.
Cook, J. O., and Spitzer, M. E. Supplementary Report: Prompting Versus Confirmation in Paired-Associate Learning 275
Cotton, J. W. See Lewis, D. J.
Curry, C. Supplementary Report: The Effect of Verbal Reinforcement Combinations on Learning in Children 434
D'Amato, M. R., and Jagoda, H. Effects of Extinction Trials on Discrimination Reversal 254
Davis, J. D. See Grice, G. R.
Davis, R. C. See Berry, R. N.
Denenberg, V. H., and Karas, G. C. Supplementary Report: The Yerkes-Dodson Law and Shift in Task Difficulty 429
DeSoto, C. B., Coleman, E. B., and Putnam, P. L. Predictions of Sequences of Successes and Failures 41
Engen, T., and Pfaffmann, C. Absolute Judgments of Odor Quality 214
Engen, T. See Miller, C.-A.
Ernst, R. L., Hoffeld, D. R., Seidenstein, S., and Brogden, W. J. Relation of Serial Position Errors to Doublet and Split-Doublet Location in Verbal Maze Pattern 94
Freeman, M. See Runquist, W. N.
Fuchs, S. S. Replication Report: An Attempt to Obtain Inhibition with Reinforcement. 343
Grant, D. A. See Beeman, E. Y.
Grant, D. A. See Hansche, W. J.
Grice, G. R., and Davis, J. D. Effect of Concurrent Responses on the Evocation and Generalization of the Conditioned Eyeblink 391

Hansche, W. J., and Grant, D. A. Onset Versus Termination of a Stimulus as the CS in Eyelid Conditioning 19
Harrison, R. H. See Jenkins, H. M.
Hartman, T. F. See Beeman, E. Y.
Hauty, G. T. See Pearson, R. G.
Helson, H., and Nash, M. C. Anchor, Contrast, and Paradoxical Distance Effects 113
Hillix, W. A. Visual Pattern Identification as a Function of Fill and Distortion 192
Hoffeld, D. R., Kendall, S. B., Thompson, R. F., and Brogden, W. J. Effect of Amount of Preconditioning Training upon the Magnitude of Sensory Preconditioning 198
Hoffeld, D. R. See Ernst, R. L.
House, B. J., and Zeaman, D. Transfer of a Discrimination from Objects to Patterns 298
Hovland, C. I. See Cahill, H. E.
Hovland, C. I. See Hunt, E. B.
Hunt, E. B., and Hovland, C. I. Order of Consideration of Different Types of Concepts 220
Indow, T., and Kanazawa, K. Multidimensional Mapping of Munsell Colors Varying in Hue, Chroma, and Value 330
Indow, T., and Uchizono, T. Multidimensional Mapping of Munsell Colors Varying in Hue and Chroma 321
Jagoda, H. See D'Amato, M. R.
Jenkins, H. M., and Harrison, R. H. Effect of Discrimination Training on Auditory Generalization 246
Jensen, G. D. Learning and Performance as Functions of Ration Size, Hours of Privation, and Effort Requirement 261
Kanazawa, K. See Indow, T.
Karas, G. C. See Denenberg, V. H.
Kausler, D. H. See Trapp, E. P.
Kendall, S. B. See Hoffeld, D. R.
Kimmel, H. D. The Relationship between Direction and Amount of Stimulus Change and Amount of Perceptual Disparity Response 68
Lewis, D. J., and Cotton, J. W. Effect of Runway Size and Drive Strength on Acquisition and Extinction 402
Lindley, R. H. Association Value and Familiarity in Serial Verbal Learning 366
Lit, A. The Magnitude of the Pulfrich Stereophenomenon as a Function of Target Velocity 165
Lloyd, K. E. Retention of Responses to Stimulus Classes and to Specific Stimuli 54
Lloyd, K. E. Supplementary Report: Retention and Transfer of Responses to Stimulus Classes 206
Lovaas, O. I. The Relationship of Induced Muscular Tension, Tension Level, and Manifest Anxiety in Learning 145
Lovaas, O. I. Supplementary Report: The Relationship of Induced Muscular Tension to Manifest Anxiety in Learning 205
Mack, J. D. See Stevens, J. C.
Marx, M. H. Resistance to Extinction as a Function of Degree of Reproduction of Training Conditions 337
Miller, C.-A., and Engen, T. Supplementary Report: Context Effects on Absolute Judgments of Length 276
Morrison, H. W. See Peak, H.
McConnell, D., and Shelly, M. W. Tracking Performance on a Sequence of Step Functions which Approaches a Continuous Function as a Limit 312
Namikas, G., and Archer, E. J. Motor Skill Transfer as a Function of Intertask Interval and Pretransfer Task Difficulty 109
Namikas, G., Thompson, C. P., and Brogden, W. J. Effect of Triplet and Quadruplicate Location in Verbal Maze Patterns upon Serial Position Errors 383
Nash, M. C. See Helson, H.
Noble, C. E. Supplementary Report: Familiarity and Frequency 432
Palola, E. G. See Sarason, I. G.
Peak, H., Morrison, H. W., and Quinn, R. P. The Generalization of Attitude Change within a Serial Structure 281
Pearson, R. G., and Hauty, G. T. Roles of Postural Experiences in Proprioceptive Perception of Verticality 425

PFAFFLIN, S. M. Stimulus Meaning in Stimulus Predifferentiation.................. 269
PFAFFMANN, C. See ENGEN, T.
PISHKIN, V. Effects of Probability of Misinformation and Number of Irrelevant Dimensions upon Concept Identification................................ 371
POLLACK, I. Identification of Visual Correlational Scatterplots...................... 351
POSTMAN, L., AND ADAMS, P. A. Studies in Incidental Learning: VIII. The Effects of Contextual Determination.................................... 153
PUTNAM, P. L. See DESOTO, C. B.
QUINN, R. P. See PEAK, H.
REMPLE, R. See WIKE, E. L.
ROBBIN, J. S. See BOTWINICK, J.
ROSS, L. W., AND SPENCE, K. W. Eyelid Conditioning Performance under Partial Reinforcement as a Function of UCS Intensity.................................. 379
RUNQUIST, W. N., AND FREEMAN, M. Roles of Association Value and Syllable Familiarization in Verbal Discrimination Learning................................ 396
RUNQUIST, W. N. See SCHULZ, R. W.
RYAN, F. J. See BILODEAU, E. A.
SARASON, I. G., AND PALOLA, E. G. The Relationship of Test and General Anxiety, Difficulty of Task, and Experimental Instructions to Performance................ 185
SCHULZ, R. W., AND RUNQUIST, W. N. Learning and Retention of Paired Adjectives as a Function of Percentage Occurrence of Response Members................. 409
SEIDENSTEIN, S. See ERNST, R. L.
SHELLY, M. W., II. Effects of Increments of Reinforcement in Human Probability Learning.. 345
SHELLY, M. W., II. Supplementary Report: Further Factors Affecting the Probability of Changing Responses...................................... 430
SHELLY, M. W., II. See MCCONNELL, D.
SILVERMAN, R. E. Eliminating a Conditioned GSR by the Reduction of Experimental Anxiety... 122
SLAMECKA, N. J. Retroactive Inhibition of Connected Discourse as a Function of Practice Level.. 104
SOLARZ, A. K. Latency of Instrumental Responses as a Function of Compatibility with the Meaning of Eliciting Verbal Signs................................ 239
SPENCE, K. W. See ROSS, L. E.
SPENCE, K. W. See TRAPOLD, M. A.
SPIKER, C. C. See WHITE, B. N.
SPITZER, M. E. See COOK, J. O.
STANDISH, R. R., AND CHAMPION, R. A. Task Difficulty and Drive in Verbal Learning.. 361
STEVENS, J. C., MACK, J. D., AND STEVENS, S. S. Growth of Sensation on Seven Continua as Measured by Force of Handgrip.................................... 60
STEVENS, S. S. See STEVENS, J. C.
THOMPSON, C. P. See NAMIKAS, G.
THOMPSON, R. F. See HOFFELD, D. R.
TRAPOLD, M. A., AND SPENCE, K. W. Performance Changes in Eyelid Conditioning as Related to the Motivational and Reinforcing Properties of the UCS............... 209
TRAPP, E. P., AND KAUSLER, D. H. Relationship between MAS Scores and Association Values of Nonsense Syllables.. 233
UCHIZONO, T. See INDOW, T.
VOSS, J. F., AND ZIEGLER, J. A. Effect of Number of Doublets upon Verbal Maze Learning.. 182
WALKER, E. G. Eyelid Conditioning as a Function of Intensity of Conditioned and Unconditioned Stimuli.. 303
WHITE, B. N., AND SPIKER, C. C. The Effect of Stimulus Similarity on Amount of Cue-Position Patterning in Discrimination Problems........................ 131
WIKE, E. L., AND REMPLE, R. Supplementary Report: Runway Performance as a Function of Reinforcement Schedule and Alley Length........................ 277
WILSON, W. A., JR. Supplementary Report: Two-Choice Behavior of Monkeys........ 207
ZEAMAN, D. See HOUSE, B. J.
ZIEGLER, J. A. See VOSS, J. F.

Journal of
Experimental Psychology

VOL. 59, NO. 1 JANUARY, 1960

TYPES OF CONFLICT AND THEIR RESOLUTION: A REINTERPRETATION[1]

JOHN M. ATTHOWE, JR.

University of Alabama

The effect of differing types of choice situations upon decision making was investigated. The subclass of choice situations examined was that of conflict. Conflict was defined as that situation of choice in which two alternatives are reacted to as approximately equal in preference. Types of conflict were defined in terms of the expected outcomes (anticipated reinforcements) associated with each alternative. Outcomes were designated as either attractive or repulsive, after Lewin (1935). In the present study, attraction or approach was equated with the possibility of gaining money. Repulsion or avoidance was equated with the possibility of losing money.

Lewin (1935) described three discrete types of conflict which he called the approach-avoidance (*ap-av*), *ap-ap*, and *av-av* types. A fourth, the double *ap-av* type, has been described by Hovland and Sears (1938). This latter type is an extension of Lewin's *ap-av* conflict. In the double *ap-av* type, S is influenced by the possible consequences of both an *ap* and an *av* outcome associated with each alternative. The weighing of four possible outcomes not only introduces a more complex set of decision operations, but also it is more nearly representative of the conflicts faced in everyday life.

The four types of conflict described by Lewin and Hovland and Sears have often been considered prototypes of conflict. The present paper views types of conflict not as discrete situations but as points along a dimension or continuum defined by the relative magnitudes of the *ap* and *av* outcomes. Position on this continuum is specified by the following ratio:

$$\frac{ap - av}{ap + av}$$

Ap and *av* refer to the total magnitudes of these tendencies within the situation of choice. This ratio is applicable only to situations of indifference (i.e., conflict). As the *ap* outcomes come to dominate, this ratio approaches the limit of +1.00. As the *av* outcomes come to dominate, the ratio approaches the limit of −1.00. At one end of the continuum lies the *ap-ap* and multiple *ap-ap* conflicts; at the other pole lies the *av-av* and multiple *av-av* types. In other words, at +1.00 *S* is sure to gain; at −1.00 he is sure to lose.[2] At the midpoint, the relative magnitudes of the *ap* and *av* outcomes (i.e., gains vs. losses) are equal; here the ratio is equal to 0.00. This is the point of "complete ambivalence." Between the midpoint and the *ap* pole lies an array of conflict types, varying only in the relative magnitude of the *ap* and the *av* outcomes, in which the *ap* outcomes dominate. Between the midpoint and the *av* pole lies an array of points representing types of conflict in which the *av* outcomes dominate.

---

[1] This article is based upon a dissertation submitted to the Department of Psychology, Stanford University in partial fulfillment of the requirements for the Ph.D. degree. The author is grateful to R. R. Sears, P. R. Farnsworth, L. Festinger, K. Little, and P. S. Siegel for their suggestions and criticisms.

Another and theoretically more parsimonious way of characterizing this dimension is in terms of risk (see Fig. 1). Risk may be defined as the relative magnitude of the total *ap* and *av* tendencies present in the situation of choice. The degree of risk will be a direct function of the above ratio. The *ap* and *av* poles (±1.00) represent the minimal risk ends of the continuum. The midpoint, 0.00, represents the point of maximal risk. In other words, types of conflict may be represented as varying degrees of risk.

Six different types of conflict, or situations of risk, were examined in this paper. They may be designated as follows: (1) the double *ap-ap* type; (2) the double *av-av* type; (3) a type consisting of an *ap-ap* and an *ap-av* alternative; (4) a type consisting of an *av-av* and an *av-ap* alternative; (5) the double *ap-av* type, where the gain was always larger than the loss; (6) the double *av-ap* type, where the loss was always larger.

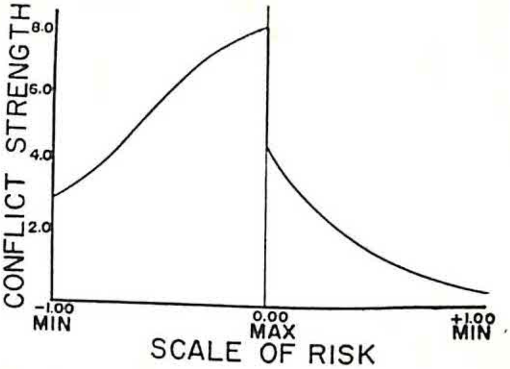


FIG. 1. Theoretical relationship between risk (type of conflict) and conflict strength.

These six types may be reduced to three major types of conflict (see Table 1). Each would be represented by an *ap*-dominant (possible gains are always greater) and an *av*-dominant (possible losses are always greater) subtype. Situations 1 and 2 are both characterized by uniformity of reinforcement associated with the outcomes. All of the outcomes are approximately equivalent for both subtypes, except for the possibility of winning in Situation 1 and losing in Situation 2. Situations 1 and 2 have been labeled the *univalent* conflict (Type I) and represent extensions of the familiar *ap-ap* and *av-av* types. Conflict Situations 3 and 4 are distinguishable by the dissimilarity of the alternatives. In both subtypes, a "gamble" is opposed to a "sure thing." The alternatives are approximately the same in each subtype but reversed in *ap* and *av* designation. Because of the dissimilarity between the competing alterna-

---

[2] As long as *S* is forced to relinquish the alternative not chosen, a pure *ap-ap* or *av-av* type conflict would not exist. There is always the possibility of making a better (*ap*) or worse (*av*) choice. If *S* knows he can simultaneously obtain both alternatives, there would be no risk present and, hence, no conflict.

TABLE 1

TYPES OF CONFLICT

(*ap* SUBTYPES ARE PRESENTED TO THE LEFT AND *av* SUBTYPES TO THE RIGHT)

	A	B		A	B
Type I. Univalence, minimal risk					
a	win 35¢	win 30¢	a	lose 34¢	lose 30¢
b	win 26¢	win x¢	b	lose 25¢	lose x¢
Type II. Disvalence					
a	win 35¢	win 5¢	a	win 26¢	lose 5¢
b	lose x¢	win 5¢	b	lose x¢	lose 5¢
Type III. Ambivalence, maximal risk					
a	win 34¢	win 26¢	a	win 25¢	win 17¢
b	lose 25¢	lose x¢	b	lose 34¢	lose x¢

Note.—A and B are alternative choices; a and b represent possible outcomes. The probability of occurrence of any a or b was equal to one-half. x represents the unique value producing indifference of choice between A and B.

tives, these two subtypes have been labeled the *disvalent* conflict (Type II). Conflict Situations 5 and 6 are subtypes of the familiar *ambivalent*, or double *ap-av*, conflict. The expected gain is always larger in Situation 5 and smaller in Situation 6. The outcomes are the same in both subtypes, but the *ap* and *av* signs are reversed. The three major conflict types are assumed to be represented along the dimension of risk such that their subtypes are symmetrically distributed about the midpoint (see Fig. 2). The greatest amount of risk would attend the ambivalent conflict and the least amount the univalent conflict.

It is generally concluded that *av* situations yield stronger and more persistent responses than equivalent *ap* situations (Atthowe, Davidson, Siegel, & Suppes, 1954; Barker, 1942; Hovland & Sears, 1938; Lewin, 1935; McClelland, 1951; Miller, 1944; Nogee, 1953). Therefore, we should expect the function representing the strength of a conflict to be higher on the *av* side of the continuum. Figure 1 represents the theoretically expected relationship between risk (the dimension underlying types of conflict) and conflict strength. The shape of the curve must take into account the difference between *ap* and *av* instigated behavior and represent a nonmonotonic function.

The present study divides into three problems, each identified with a measure of conflict resolution. The first problem is concerned with the effect of differing degrees of risk upon the latency of a response. Response latency was defined as the time it takes *S* to make a decision. Decision time has been shown to be one of the most "sensitive" measures of choice or judgment (Cartwright, 1941). In general, as two objects or events approach "equivalence," the longer becomes the decision time (Deutsch, 1954; Johnson, 1945). Decision time has also been related to the degree of confidence in one's decision (Johnson, 1945) and to emotional behavior (e.g., complex detection [Woodworth, 1938]).

It is hypothesized that decision time as an index of conflict strength will be related to risk, or type of conflict, as represented by the theoretical curve in Fig. 1. The following predictions would, thereby, be generated:

Hypothesis I. The three major types of conflict will differ significantly in decision time.

Corollary 1a. The ambivalent conflict will show the longest decision times.

Corollary 1b. The univalent conflict will show the shortest decision times.

Hypothesis II. The *av*-dominant subtypes will show longer decision times than the *ap*-dominant subtypes.

In addition, we are concerned with the effect of risk upon the frequency and degree of nonrational responses. Choices were designated as either rational or nonrational. A choice equivalent to that made by a mathematical model, or a purely "rational man," was scored as rational; any choice differing from this choice was termed nonrational. The larger the difference between the observed choice and rational choice, the greater would be the degree of nonrationality. The meaning of a rational or nonrational response is unclear. Any definition must at present be arbitrary; nevertheless, a criterion of "maximization" is probably essential. Nonrational, as well as rational, behavior is certainly found in situations involving conflict, but there is little evidence relating degree of nonrationality to the type of conflict. Is nonrationality a measure of conflict strength?

The following general questions were asked in regard to the rationality or nonrationality of a decision conflict: (1) What is the relationship between risk (type of conflict) and the degree of rationality of a decision? (2) What is the relationship between rationality and decision time?

The third problem dealt with *S*s' strategy, or pattern of responses, and the relationship of strategy to risk. Two general types of nonrational strategies were identified together with the "completely rational" strategy. The two nonrational strategies were: (*a*) *conservatism*, an overconcern with the possibility of losing money, and (*b*) *extravagance*, an overconcern with the possibility of gaining money. An *S* who more frequently chose the alternative representing the least possibility of losing rather than the alternative representing the greatest possibility of winning was termed a conservative strategist. If the latter alternative was chosen more frequently, *S* was labeled extravagant. For example, in deciding between Alternatives A and B below, a conservative would choose B, and an extravagant *S* would choose A.

	A	B	
(*a*)	win 35¢	win 26¢	$P = \frac{1}{2}$
(*b*)	lose 26¢	lose 17¢	$P = \frac{1}{2}$

A completely rational *S* would indicate that both alternatives were equivalent in expected value.

Game theorists (e.g., Dunlap, 1955; Luce & Raiffa, 1957; Thrall, Coombs, & Davis, 1954) have introduced the concept of strategy into experimental psychology. Their emphasis has been upon the best type of strategy. The emphasis in this paper is upon the types of strategy typically used and the conditions associated with a given strategy. Specifically, the questions asked were the following: (1) What are the typical types of strategies used in resolving conflicts? (2) What is the relationship between risk and the strategy used?

METHOD

Task.—A simple yet fairly representative task involving a series of choices was developed. The task, a gambling game, was both interesting and motivating. The instructions were designed to induce moderate to strong achievement motivation and a reasoning attitude. The alternative choices were represented as "wagers." They were said to be mathematical reasoning problems taken from an advanced intelligence test. The *S*s were told that the objective of the investigation was to study the influence of a gaming context upon general reasoning ability.

The *S*s were instructed to choose one of two alternative wagers. Each alternative consisted of two possible outcomes (see Table 1). The probability of either outcome was one-half. The outcomes symbolically represented differing monetary values, although no money was actually involved. In making his choice, *S* had to weigh the possible numerical consequences (i.e., outcomes) of each alternative before making his decision. Because no money was gained or lost and *S*s were never told whether their decisions were correct, the effects of reinforcement were assumed to be nullified.

The event determining the occurrence of either outcome was stated to be the rolling of a die. If the die rolled an even number, Outcome a was said to occur (see Table 1); otherwise Outcome b occurred. The actual rolling of the die was never performed. The die was simply used as a prop.

Subjects.—The *S*s were 32 male students attending a class in introductory psychology at Stanford University. They were each tested individually for approximately 1.5 hr.

Procedure.—Each of the six conflict types was randomly presented in the series of choices. In each type, the decision analyzed was that made at the point of conflict. The point of conflict was taken as the point of indifference between the two alternatives (i.e., the PSE). The indifference point was established by the method of constant stimuli. The outcomes of one alternative (e.g., Alternative A in the *ap* subtype, ambivalent conflict, Table 1) were held constant throughout the session; whereas one of the outcomes (Outcome b in Alternative B) of the other alternative was varied. This outcome, designated by x, was varied until the two alternatives were reacted to as indifferent. For example, if x in the above subtype assumed at random the values 10, 14, 16, 18, 20, and 24, and *S*'s choices were B, B, B, A, A, and A, then the point of indifference lies midway in the interval bounded by 16 and 18 (i.e., 17).

The time required to make a choice and the particular choice were both determined at the point of indifference. Decision time was measured in units of .1 sec. with an electric timer. A rational, or mathematically correct, choice was operationally defined as that choice in which the mathematical expected values of the alternatives were equal in magnitude. In the above example, 17 is also the rational choice. Any choice corresponding to an indifference point which deviated from this value was termed nonrational. In the above example, if *S* had preferred B for the values of 10, 14, 16, and 18 and A for 20 and 24, then his indifference point would be 19. In this case, the degree of nonrationality represented by 19 would be 2 (i.e., 2 cents or units away from a rational choice).

The *S* was separated from *E* by a low screen. The wagers, printed on 3 × 5 cards, were presented to *S* through a small opening at the bottom of the screen. As the card emerged on *S*'s side of the screen, the timer was started. The *S* was instructed to make only one choice. An *O*, seated a few feet away, recorded choices and decision times. The *E* also kept a tally of *S*'s choices in order to determine *S*'s point of indifference. If 12 choices were made within a given subtype without the determination of an indifference point, *S* was automatically discarded. Seven *S*s were discarded for this reason.

RESULTS

Decision time.—The distribution of decision times was markedly skewed. Further, Bartlett's test (Edwards, 1950) for homogeneity of variance was rejected at the .001 level. Neither logarithmic nor reciprocal transformations increased the homogeneity of variance. Therefore, ordinal measurement was assumed, and distribution-free methods were used to analyze the data.

Hypothesis I was confirmed. The six subtypes as well as the three major types of conflict yielded significantly different decision times ($P \leq .001$ in both instances). The test of significance used was Friedman's two-way analysis of variance of ranked data (Siegel, 1956).

Corollary 1a was confirmed. The difference between the ambivalent and disvalent types (the smallest difference) was determined for each *S* and then ranked in magnitude. The ambivalent type of conflict yielded significantly longer decision times ($P \leq .025$). The Wilcoxon Signed-Ranks test (Siegel, 1956) was used to test the difference.

TABLE 2

Conflict Resolution as a Function of Type of Conflict

Type of Conflict	Nonrationality		Decision Times (Sec.)	
	Freq. $N = 26$	Degree $N = 19$	Mean $N = 19$	Rank $N = 26$
Ambivalence				
ap	17	36	21.8	3.3
av	15	20	27.9	5.1
Disvalence				
ap	23	135	20.7	3.0
av	25	183	25.0	4.2
Univalence				
ap	12	30	18.1	2.6
av	10	16	19.8	2.9

Note.—Rank represents the mean of the rankings for a given type of conflict. A rank of 6.0 equals the longest decision time.

Corollary 1b was confirmed. The difference between the univalent and disvalent types (the smallest difference) was determined. The univalent conflict yielded significantly shorter decision times ($P \leq .05$, Wilcoxon test).

Hypothesis II was confirmed. The *av* subtypes showed significantly longer decision times than the *ap* subtypes ($P \leq .01$, Wilcoxon test).

The results are presented in tabular form in Table 2. Six *S*s showed a rational strategy (i.e., showed no nonrational choices). Consequently, they did not enter into the computations; nevertheless, their responses as a group served to clarify the relationship between risk and decision time. The six rational *S*s also showed significant differences in decision time as a function of the type of conflict ($P \leq .05$, Friedman test, $df = 5$). The ambivalent conflict yielded the longest decision times, but the disvalent conflict showed the shortest decision times. All statistical tests reported in this section were one-tailed tests, except the latter.

Nonrationality.—Both the frequency and the degree of nonrational choices were found to be significantly different among the subtypes and between the three major types of conflict ($P \leq .001$). Cochran's Q test (Siegel, 1956) was used to evaluate frequency and Friedman's test to evaluate degree. The disvalent conflict clearly yielded the largest frequency (Q test) and the greatest degree (Wilcoxon test) of nonrationality when compared to either of the other major types ($P \leq .001$). The univalent conflict generally exhibited the fewest number (Q test) and the least degree (Mood's Chi-square test for correlated measures [McNemar, 1955]); the latter trends approached but did not reach significance ($P \leq .15$ in both instances). Because the differences were often similar in magnitude, Mood's test was considered more appropriate than that of Wilcoxon.

No significant difference was shown between the *av* and the equivalent *ap* subtypes, either in terms of the frequency or the degree of nonrationality. The rank-order correlation (rho) between the degree of nonrationality exhibited by *S* and his median decision time was .04. This independence between the rationality and the latency

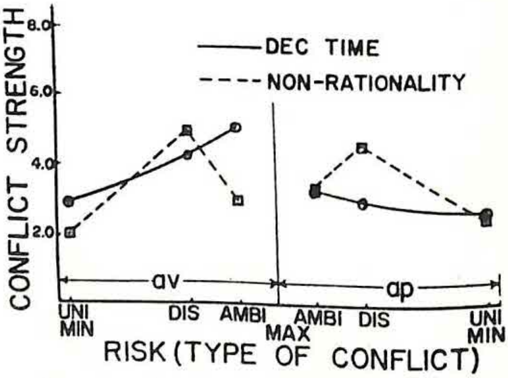


FIG. 2. Conflict strength as a function of risk. One scalar unit of strength represents one unit of decision time (ranks) and five units of nonrationality (frequency).

of a decision is further shown in Fig. 2. Whereas the relationship between risk and decision time is monotonically increasing, the relationship between risk and nonrationality was nonmonotonic. All statistical tests involving nonrationality were two-tailed tests.

Strategy.—Of 18 *S*s who were classified as having dominant nonrational strategies, 83% were conservative. This difference was significantly different from chance expectations at the .01 level, as determined by chi square. If all nonrational choices are considered, 71% were in a conservative direction. As would be expected, all 13 *S*s who consistently refused to take a chance (i.e., gamble) in the disvalent conflict used a conservative strategy ($P \leq .001$, chi square, when compared with chance expectations). The one *S* who consistently took a chance used an extravagant strategy. Four *S*s varied their strategies from situation to situation.

An inspection of Table 3 indicates that more rational decisions were made in the univalent situation; no distinction was made between the nonrational strategies. The strategy exhibited in the disvalent conflict was predominantly conservative. In the ambivalent conflict, very few decisions were extravagant; both conservative and rational strategies were more frequently manifested. If the total *N* is inflated by assuming that the observations in each choice situation were made independently, a chi square can be computed. The overall and pairwise tests of significance (two-tailed) were beyond the .001 and .01 levels when the three types of conflicts were compared in terms strategy.

TABLE 3

STRATEGY AS A FUNCTION OF TYPE OF CONFLICT

Strategy	Type of Conflict						Total
	Ambivalent		Disvalent		Univalent		
	Av	*Ap*	*Av*	*Ap*	*Av*	*Ap*	
Conservative	10	14	20	17	4	7	72
Extravagant	5	3	5	6	6	5	30
Rational	17	15	7	9	22	20	90
Total	32	32	32	32	32	32	192

Note.—The tabled values refer to the number of *S*s exhibiting a given strategy.

DISCUSSION

The results indicate the necessity of considering the type of conflict, or risk, in predicting decision making behavior. By equating the outcomes along four dimensions, differences in decision making were assumed to be a function of the situational arrangement of the outcomes. The subjective probability of the occurrence of any outcome, the relative magnitude of their difference within an alternative, the absolute expected value (mathematical) of the sum of the outcomes per alternative, and the range of outcome values were held constant or relatively so. Table 1 shows that all of these controls could not be carried out simultaneously. The expected values of the univalent alternatives were approximately equal to 30; those of the other alternatives were approximately 5. The net effect of this discrepancy would be to overweight the univalent conflict and reduce the likelihood of reaching significance; consequently, the substantiation of the hypotheses may be viewed with greater certainty.

The relative balance of *ap* and *av* tendencies (i.e., risk) was shown to be directly related to the strength of a conflict as measured by decision time. The greater the risk, the longer the decision time. The relationship between risk and nonrationality was not as clear. The results indicate that a moderate amount of risk was associated with maximum nonrationality, although the conflict representing moderate risk (disvalence) was also characterized by the option of gambling. It seems more likely that the instrumental uncertainty (whether to gamble or not) involved in making a decision is more important in determining its correctness. This latter interpretation is suggested by the comparison of the completely "rational" with the nonrational *S*s. Both groups exhibited longer decision times in the ambivalent conflict, but the rational *S*s' shortest latencies were in the disvalent conflict, while those of the nonrational *S*s were in the univalent conflict. Risk alone cannot account for this difference. It is suggested that a concomitant of rationality is a reduction in instrumental uncertainty, in this case the concern with taking chances.

The *S*s tended to minimize their maximum losses rather than maximizing their maximum gain. These findings lend empirical support to the "minimax" principles found in game theory (e.g., Luce & Raiffa, 1957); although, as Mosteller and Nogee (1951) have shown, college students are probably a rather conservative sample. In addition, the type of conflict or risk provides a basis for differences in strategy. The *S*s tended to be significantly more rational when risk and uncertainty were minimal. As risk and instrumental uncertainty increased, conservatism also increased. The option of gambling not only was associated with more nonrational decisions but also with much greater conservatism.

The equation of types of conflict and risk provides a basis for a clearer and more formal exposition of conflict theory. The focus upon anticipated outcomes also suggests an interrelationship between decision making (judgmental processes) and conflict (motivational processes).

The gaming paradigm has a far wider experimental application than in the testing of statistical models, per se. Its appeal to many diverse populations (e.g., Hurst & Siegel, 1956; Mosteller & Nogee, 1951) plus its simplicity provides a ready experimental context in which to test and to compare judgmental and motivational conditions. In addition, the method of indifference-point analysis yields a basis for equating and comparing individual reactions and situations of conflict.

SUMMARY

The relationship between decision-making behavior and differing types of conflict was investigated. The experimental task involved choosing the "better" of two alternative wagers. Conflict was defined in terms of the arrangement of the expected outcomes. Risk, the relative magnitudes of *ap* and *av* outcomes, was postulated to be one of the major dimensions underlying types of conflict. As risk increased, the conflict strength, as defined by decision time, also increased. Rationality was associated with minimal risk, whereas nonrationality was mainly associated with uncertainty (chance taking). Decision time and the rationality of a decision were found to be independent measures of conflict resolution. The *S*s tended to be more conservative, especially as risk and uncertainty increased.

REFERENCES

ATTHOWE, J. M., JR., DAVIDSON, D., SIEGEL, S., & SUPPES, P. An experimental study of some factors influencing decision-making in conflict situations involving risk. Paper read at West. Psychol. Ass., Long Beach, April, 1954.

BARKER, R. G. An experimental study of the resolution of conflict by children. In Q. McNemar & M. K. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942. Pp. 13–34.

CARTWRIGHT, D. Decision-time in relation to the differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442.

DEUTSCH, M. Field theory in social psychology. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 181–222.

DUNLAP, J. W. *Mathematical models for human behavior*. Stamford, Conn.: Dunlap Assoc., 1955.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

HOVLAND, C. I., & SEARS, R. R. Experiments on motor conflict: I. Types of conflict and their mode of resolution. *J. exp. Psychol.*, 1938, **23,** 477–493.

Hurst, P. M., & Siegel, S. Prediction of decisions from a higher ordered metric scale of utility. *J. exp. Psychol.*, 1956, **52,** 138–144.

Johnson, D. M. A systematic treatment of judgment. *Psychol. Bull.*, 1945, **42,** 193–224.

Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

Luce, R. D., & Raiffa, H. *Games and decisions: Introduction and critical survey.* New York: Wiley, 1957.

McClelland, D. C. *Personality.* New York: Sloane, 1951.

McNemar, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. I. New York: Ronald, 1944. Pp. 431–465.

Mosteller, F., & Nogee, P. An experimental measurement of utility. *J. polit. Econ.*, 1951, **59,** 371–404.

Nogee, P. An experiment in decision-making under uncertainty. Paper read at the Seminar in the Application of Mathematics to the Social Sciences. Univer. of Michigan, November, 1953.

Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

Thrall, R. M., Coombs, C. H., & Davis, R. L. (Eds.), *Decision processes.* New York: Wiley, 1954.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 3, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

AGE DIFFERENCES IN CARD-SORTING PERFORMANCE IN RELATION TO TASK DIFFICULTY, TASK SET, AND PRACTICE

JACK BOTWINICK, JOSEPH S. ROBBIN, AND JOSEPH F. BRINLEY[1]

Laboratory of Psychology, National Institute of Mental Health

Inferences about aging processes may be made from functional relationships between task difficulty and age differences in task performance. For example, from such relationships, Welford (1958) has suggested that information capacity of the sensorimotor system is lowered with age, Botwinick, Brinley, and Robbin (1958a) have suggested that with age there may be an increase in response reviewing tendencies, and Birren and Botwinick (1955) have emphasized perceptual changes with age.

In the present study, performances of two age groups were compared when task difficulty was defined in two different ways but when the task itself, card sorting, was kept the same. The problem of task set was also studied. These three aspects of the task were investigated with respect to age differences in the effects of practice.

METHOD

Subjects.—There were two age groups of male volunteer *S*s. The older group comprised 27 *S*s aged 65 to 81 yr. with a median of 71.0 yr., and the younger group comprised 30 *S*s aged 19 to 35 yr. with a median of 24.5 yr.

Both age groups were divided into high- and low-education subgroups—those with 12 or more years of formal education and those with less than 12 years. In the older group, 14 *S*s were in the high-education subgroup and 13 were in the low- subgroup. The range of education was 4 to 11 and 12 to 16 yr. in low- and high- subgroups respectively. All elderly *S*s, except one, were community residing.

The high-education subgroup of young *S*s comprised 10 religious research volunteers and 8 Public Health Service (PHS) employees. The low- subgroup of younger *S*s comprised 12 workers of the PHS. The range of education was 6 to 11, and 12 to 18 yr. in low- and high- subgroups, respectively.

Procedure.—Each *S* performed in a series of three experiments that involved card-sorting tasks. The tasks were completed during the course of three sessions which lasted a total of approximately 3.5 hr. Two younger *S*s, however, completed the tasks in fewer sessions. The *S*s held 71 playing cards face up and sorted as quickly as possible into slots according to instructions. The 71 cards were of two complete decks from which the 24 face cards and the 9 stimulus cards shown in Table 1 were removed. Each card was sorted into a 3.6 × 5.0-in. slot of a slightly tilted sorting board 36 in. in the left to right dimension, divided equally into 10 slots. When a card was sorted, it slid out of sight through a narrow opening at the back end of the slot. Placed perpendicular to the slot and facing *S*, was the stimulus card that indicated the sort.

Experiment I.—Each *S* sorted the 71 cards by matching the number on the card with the number on the stimulus card. When no match was possible because none of the stimulus numbers corresponded to the sorting card, it was placed in the tenth slot, that is the last slot, which was designated "bin."

There were five levels of card-sorting difficulty defined by the number of stimulus cards or slots that were sorted. The difficulty levels were 2, 4, 6, 8, and 10 slots. Since a different number of cards were sorted into the bin for each level of difficulty, the abscissa in Fig. 1 is labeled a, b, c, d, and e, representing lesser to greater task difficulty. As indicated in Table 1, all 36 in. of the sorting board were used for each level of difficulty. Each of the five different levels was sorted four consecutive times, i.e., 4 trials. There was a shuffling of the cards between each trial. Difficulty levels were presented in random order except that Level c was always last.

---

[1] We thank James E. Birren, Chief, Section on Aging, NIMH, for suggestions relating to this study.

TABLE 1

ARRANGEMENT OF STIMULUS CARDS ON SORTING BOARD FOR VARIOUS TASK DIFFICULTY LEVELS

Task	Sorting Board Slot Numbers									
	1	2	3	4	5	6	7	8	9	10 (bin)
Warm up	Q			J			K			X
Level a	3C									X
Level b	7C			4D			9C			X
Level c	5H		9C		3C		8S		6S	X
Level d	7C		4D	10S	6S	3C	5H	2D		X
Level e	4D	10S	6S	9C	2D	5H	7C	3C	8S	X

Note.—Stimulus cards are shown in the position they occupied on the sorting board for the various levels of task difficulty. Thus, for difficulty Level a, Stimulus Card 3C occupied the first slot while for Level c, it occupied the fifth slot. All sort cards that could not be placed properly in Slots 1 through 9, were placed in tenth slot (bin). Q, J, K = face cards, C = clubs, D = diamonds, H = hearts, S = spades, and X = extra sort cards.

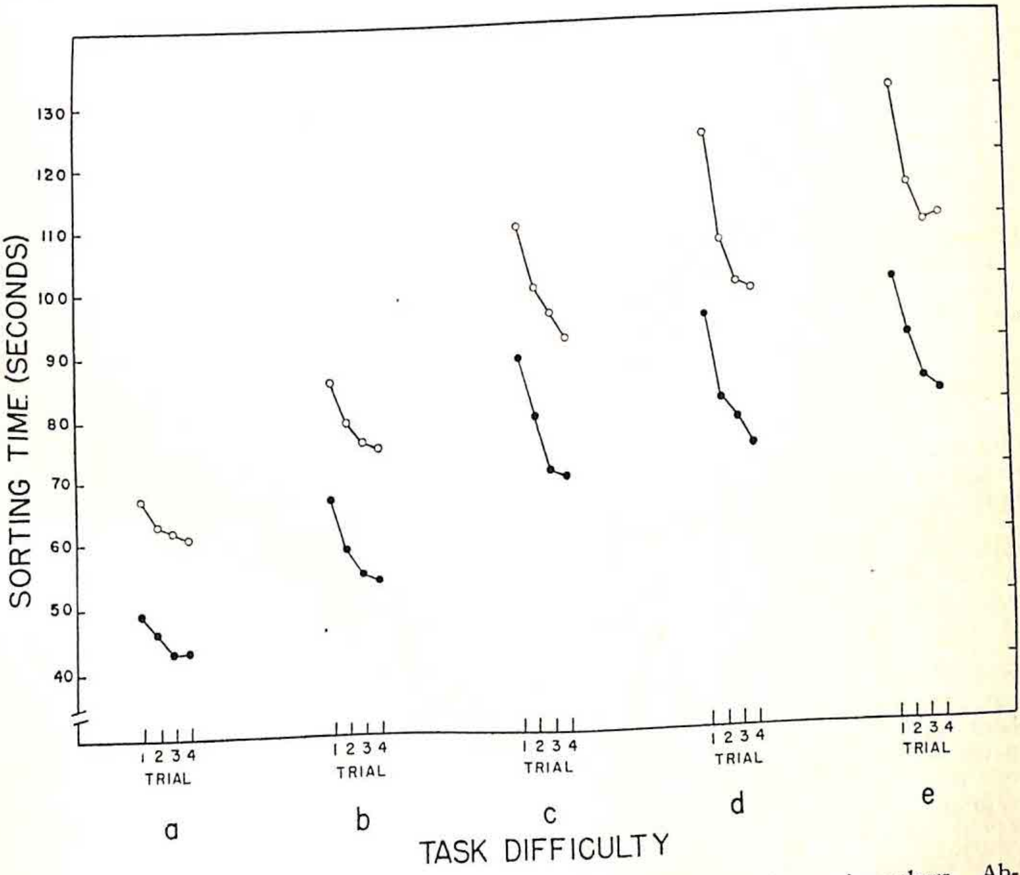


FIG. 1. Mean time to sort 71 cards by matching sort and stimulus card numbers. Abscissa refers to number of slots or match numbers, with difficulty Level a being the fewest and Level e being the largest number of slots. Each difficulty level was sorted for four trials by *S*s 65 to 81 yr. (open circles), and by *S*s 19 to 35 yr. (filled circles).

This was for a purpose that will be described later.

All *S*s were given warm-up practice before the experiment began. The warm-up involved three trials of matching face cards shown in Table 1. Each trial comprised 21 face cards, plus 8 sevens and 8 aces for sorting in bin.

Experiment II.—This experiment included an analysis of age differences in the effects of changing the task set or card-sorting habit. Immediately following the four trials of sorting with the 6-slot difficulty level (Level c), an additional four trials were given followed by a different arrangement of the same stimulus cards. The new arrangement was 9C, 6S, 8S, 5H, and 3C, and bin. Four consecutive trials were given with the new stimulus arrangement. The *S*s for this experiment were all those of the young group, 6 elderly of the high-education subgroup, and 10 elderly of the low- subgroup plus 3 additional elderly *S*s of the low- subgroup who did not perform in Exp. III.

Experiment III.—This experiment included an analysis of age differences in performance as it is related to complexity of task. The term *complexity* refers to a type of task variation which is to be distinguished from *difficulty*. Difficulty was defined by the number of slots required for a particular task. Complexity of task is defined by the number of stimulus aspects that need to be considered at one time. Most often, Exp. III was begun on a different day.

The *S*s sorted according to two different types of instructions. In Exp. I, the instructions were simply to sort by matching number of sorting card to number of stimulus card, and if this was not possible, to place the sorting card in the bin. This will be referred to as Complexity A. In Exp. III, Complexity B required the same sort "only if the color is the same; if the color is different, then place in bin." Color referred to whether the suit was red or black. Complexity C required *S*s to sort according to three stimulus aspects: number, color, and whether the numbers were odd or even. Thus, instructions were to match "cards by number and color if the number is even, but if the number is odd, then match number and opposite color." Nonmatchable cards were placed in bin. Card-sorting errors were treated in exactly the same way as were correct sorts.

Sorting was not begun until both *S* and *E* were satisfied that the instructions were understood. Instructions were repeated before each trial and as often as necessary. Within Complexities B and C, there were two levels of task difficulty: 4 and 10 slots (Levels b and e). Orders of Complexity B and C, and Difficulty Levels b and e were given for four trials each in random order. The same warm-up for sorting was given before Tasks B and C as was given before Task A. Tasks B and C were given most often in two sessions and nearly always in two separate days. In the 4-slot difficulty level for both Complexities B and C, the stimulus card 9H was substituted for the 9C on alternate trials in order to discourage correct sorting by a different method that was found to be present with early *S*s. Three elderly *S*s of the low-education subgroup did not perform in Exp. II Their scores, however, were not different from the mean of the subgroup.

RESULTS

Experiment I

The data are presented as Fig. 1. The effects of the education subgroups are not shown, but they were considered in the analyses of the data. Except for sorting trials, the data were subjected to a multivariate statistical analysis as described by Anderson (1958).[2] The effects of sorting trials

---

[2] The multivariate analysis was suggested and directed by Seymour Geisser, Section on Mathematics and Theoretical Statistics, NIMH. We are very grateful to him for his kind service. The design of the study and interpretation of results, however, are the responsibility of the authors.

The analysis involves the computation of the means, variances, and correlations of all the conditions of task difficulty in Exp. I and task complexity in Exp. III which are then combined in a manner requisite to the hypothesis. This leads to Hotelling's T^2 statistic which is transformed into an F ratio with corresponding P values. Degrees of freedom are a function of the number of conditions of difficulty or complexity, number of groups, and number of individuals in each group.

The current data require a multivariate analysis rather than an analysis of variance because the latter method assumes that the variation of individuals be the same among the different conditions of difficulty or complexity, and that the conditions be uncorrelated. These assumptions are rarely met when replicate observations are made on the

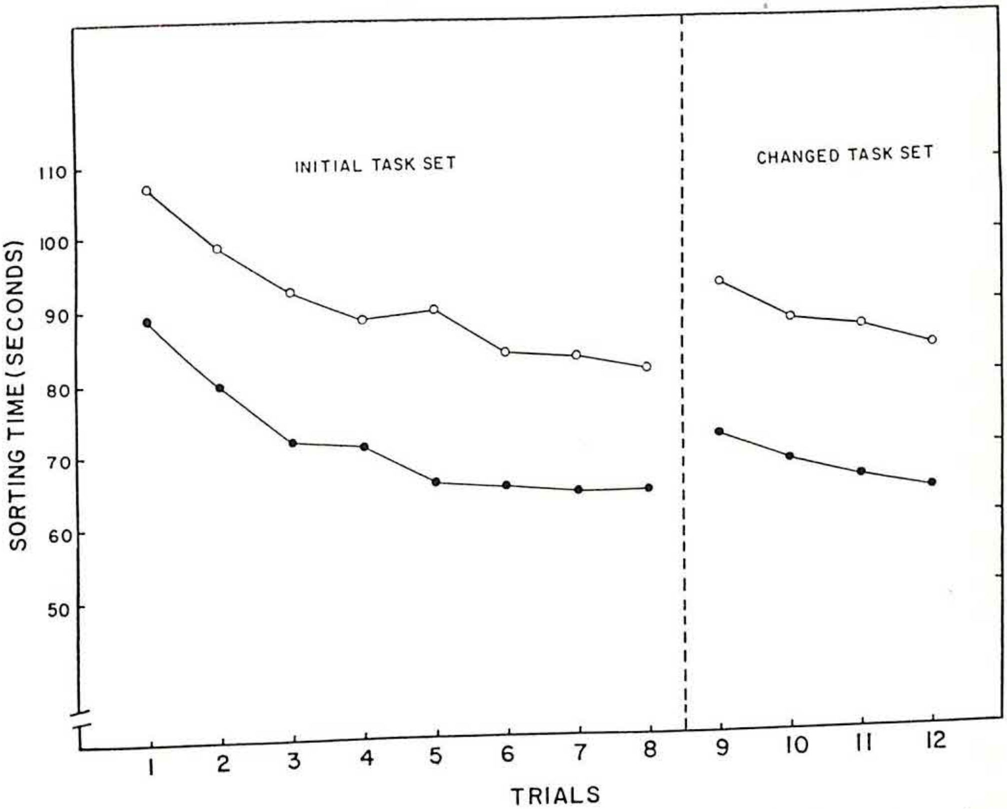


FIG. 2. Mean time to sort 71 cards by matching sort and stimulus card numbers. Each sort was done for eight consecutive trials followed by an additional four trials with a rearranged stimulus pattern. The *S*s were 65 to 81 yr. (open circles), and 19 to 35 yr. (filled circles).

were analyzed by *t* tests utilizing the values associated only with Trials 1 and 4.[3]

The older group was found to sort slower than the younger group, and sorting speed decreased in relation to increased task difficulty. There was improvement in speed of sorting from Trial 1 to Trial 4, and this tended, in general, to occur at a faster rate with the more difficult tasks. The two education subgroups were not significantly different in their sorting speeds except in relation to the interaction of age and task difficulty. This relation (Education × Age × Task difficulty) was unsystematic and difficult to interpret, but the low-education subgroup of younger *S*s was consistently poorer than the better educated *S*s of the same age. The *F* ratios which refer to these results were statistically significant ($P < .01$), but all other tests of the analyses were not ($P > .05$).

Experiment II

The data of this experiment are presented as Fig. 2. It may be seen that with the change of task set or

---

same *S* in all treatments, i.e., levels of task difficulty or complexity.

[3] *t* tests were performed when the requirements of multivariate analysis were not met in the data or when only two conditions were involved. Variance analyses were not performed because of disproportionate subgroup frequencies.

card-sorting habit in Trial 9, both age groups required more time to sort the 71 cards. Of the 30 younger and the 19 older *S*s, all but 3 in each age group required additional time.

From Fig. 2 it may be computed that the effect of changing the stimulus pattern for the older group was approximately 1.5 times that for the younger group. The older group required, on the average, 11.9 sec. additional to sort the new pattern, and the younger group required an additional 8.0 sec. However, when the sorting time of Trial 8 was subtracted from that of Trial 9 for each *S*, and age groups compared in this difference in sorting time, it was found that no reliable difference existed ($P > .05$).

It may be seen in Fig. 2 that both age groups improved sorting performances with increased practice trials. No appreciable age difference in the rate of improvement was found, however. For each *S*, sorting time

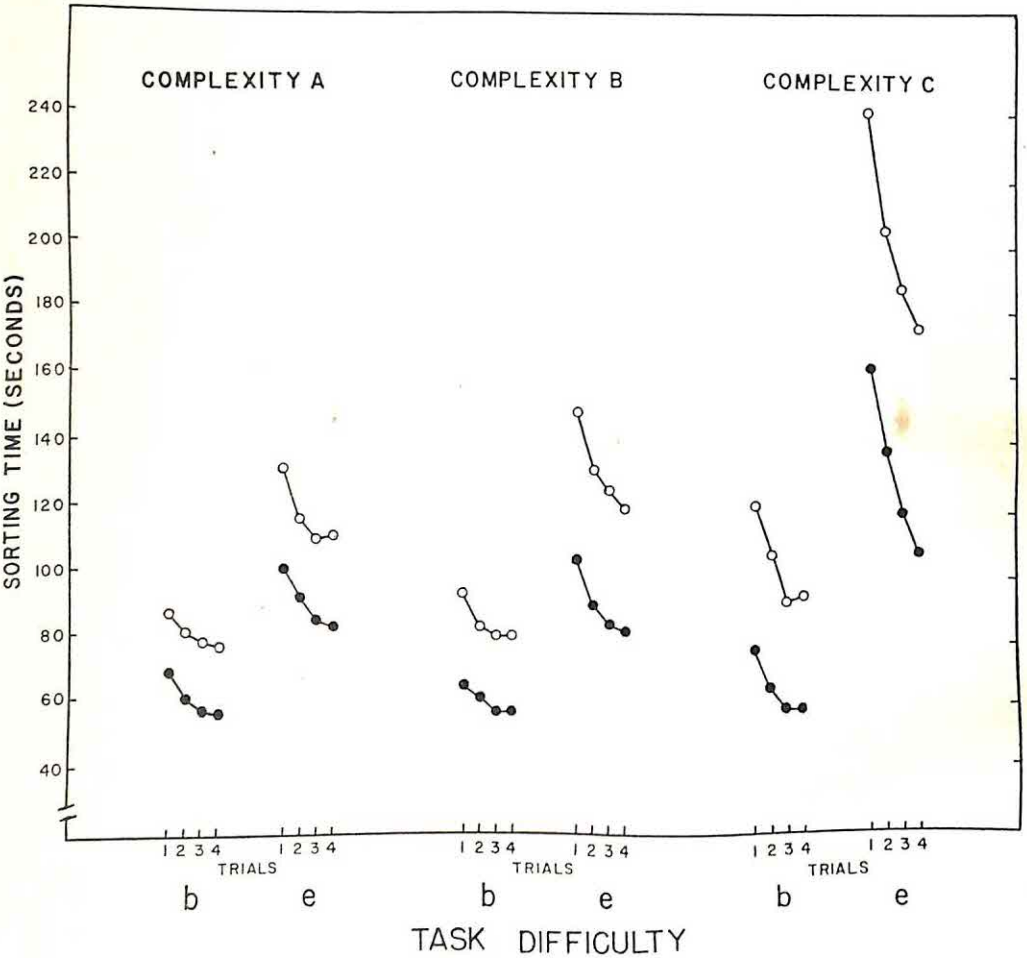


FIG. 3. Time to sort 71 cards in relation to task difficulty (number of stimulus cards matched), and task complexity (number of stimulus aspects kept in mind and manipulated simultaneously). There were four trials for each level of task complexity and difficulty. Task Level A and b involved the fewest number of stimulus cards and stimulus aspects, and Task Level C and e the largest number. The *S*s were 65 to 81 yr. (open circles), and 19 to 35 yr. (filled circles).

TABLE 2

MULTIVARIATE ANALYSIS OF DATA FROM EXP. III

Source	df^a	T^2	F
Age (A)	3, 49	38.76	12.43**
Education (E)	3, 49	2.43	.78
Task complexity (C)	2, 52	10.64	5.22**
A × C	2, 52	10.46	5.14**
A × E	3, 49	1.59	.51
E × C	2, 52	1.71	.84
Task diffiiculty (D) × C	2, 52	131.61	64.56**
Sorting trials (T) × C	2, 52	29.08	14.27**
A × E × C	2, 52	.42	2.06
A × C × D	2, 52	9.73	4.77*
A × C × T	2, 52	2.31	1.13
D × C × E	2, 52	.08	.04
D × C × T	2, 52	11.19	5.49**
T × C × E	2, 52	.97	.47
A × D × C × T	2, 52	.08	.04
A × C × T × E	2, 52	1.10	.54
A × D × C × E	2, 52	3.42	1.68
D × C × T × E	2, 52	.25	.12
A × D × C × E × T	2, 52	6.35	3.12

[a] The df's are not determined in the same way as in analysis of variance. They are based upon different statistical operations.

* $P < .05$.

** $P < .01$.

of Trial 8 was subtracted from Trial 1, and also sorting time of Trial 12 was subtracted from Trial 9, in order to obtain measures of improvement in both the original sorting task and in the task of changed stimulus pattern. No significant differences between age groups were found with respect to these measures ($P > .05$).

Experiment III

Card-sorting time was analyzed in relation to age, education, task difficulty, task complexity, sorting trials, and the interactions among these conditions. The data, except for the effects of education, may be seen in Fig. 3. Relations were evaluated statistically by multivariate and t test analyses, as before, and the results of these tests are shown in Tables 2 and 3.

The multivariate analysis indicated that the effect of increased age and increased task complexity was to increase sorting time. In addition, relative to the sorting times with the simple task, the older group was slower with increased complexity of task than was the younger group. All three differences (age, task com-

TABLE

COMPARISONS OF MEANS BY t TEST OF DATA FROM EXP. III

Source[a]	t
Trials (T)	16.48**
Task difficulty (D)	4.44**
Age (A) × T	1.51
A × D	3.79**
T × D	2.06*
Education (E) × T	1.46
E × D	.46
A × T × E	1.40
A × D × E	.77
A × D × T	.50
D × T × E	1.09
A × D × T × E	1.96

[a] df's for all sources = 53 (the sum of the four subgroup df's). Analyses of variance not performed because of disproportionate subgroups Ns.

* $P < .05$.

** $P < .01$.

plexity, and the interaction between age and task complexity) were significant ($P < .01$). It may be seen in Fig. 3 that sorting time was longest with the task that was a combination of most difficult and most complex ($P < .01$). This effect of combined task difficulty and complexity was disproportionately greater for the older group ($P < .03$).

The effects of sorting trials were investigated by use of Trials 1 and 4. Rate of improvement with practice was positively related to increased task complexity and to the interaction of increased task complexity and difficulty ($P < .01$). The data involving Sorting Trials 1 and 4 as related to age, education, and task difficulty were subjected to a series of t tests. The effect of practice was to decrease sorting time, the effect of increased task difficulty was to increase sorting time, and the improvement with practice occurred at a faster rate with the more difficult problems. Unlike the statistical results of Exp. I, the Age × Difficulty interaction was significant, indicating that the effect of increased task difficulty was to increase sorting time disproportionately for the older group. These differences, except for the Trials × Difficulty interaction, were all significant at $P < .01$; the Trials × Difficulty interaction was significant at $P < .05$. As with task-complexity, the effect of trials was similar for both age and education groups, and all the second-order interactions involving education were nonsignificant ($P > .05$).

DISCUSSION

There are three main aspects to the results of these studies. The first is that increased task difficulty and complexity tends to slow the elderly more than the young. The second aspect is that while there is improvement with continued sorting, and while rate of improvement tends to increase with increased task difficulty and complexity, improvement rate is about the same for the elderly as for the young *Ss*. The third aspect is that high- and low-education groups were not significantly different in sorting speed.

It is important to evaluate the task variations within the three experiments with respect to age. The motor or movement component of the various card-sorting performances was more or less constant from one experimental condition to another. Thus the effects of task difficulty (Exp. I), task set (Exp. II), and task complexity (Exp. III) were primarily on nonmotor aspects of card sorting. The increase in sorting time with increased task difficulty was due, therefore, more to the perceptual operations of searching for or matching of the specific card number, or to the selection and programming of responses, than to motor speed. The effect of age on this function was not appreciable. The effect of age was more apparent when the task was varied in complexity by the number of stimulus aspects that needed to be kept in mind and manipulated simultaneously. The relative increase in sorting time by the elderly with increased complexity was taken as an index of age deficit in this type of cognitive function rather than in motor speed. The number of stimulus aspects kept in mind and manipulated may be related to memory, or it may be related to the amount of material that can be integrated and kept available at any given period of time.

Crossman and Szafran (1956) also found that age was in different relation to different types of task variations. They found age differences in speed of response greater on the easy than the difficult levels of both a card-sorting and a weight-lifting task. However, with a different card-sorting task, similar in some respects to that of Exp. I in the present study, they found an increase in age difference in sorting speed with the first step of task difficulty, and then a constant age difference with further difficulty. Their results, especially when viewed in the context of the present and other studies (e.g., Birren & Botwinick, 1955; Clay, 1954; Kay, 1955), suggest that a general theory of aging, as for example, changes in channel capacity, or response reviewing tendencies, may be appropriate only to a specific class of tasks. It may be necessary to categorize tasks into classes by specific characteristics and to develop theories of aging for each class of tasks independently.

The change of task instructions or requirements, examined in Exp. II, indicated that both age groups were statistically similar with respect to the effect of change in task set. Interpretation of this finding however, may need to be qualified because of two reasons. First, change of task set involves modifiability of behavior and in recent studies, age differences in modifiability were found (Botwinick, Brinley, & Robbin, 1958b; Botwinick, Brinley, & Robbin, 1959; Botwinick, Robbin, & Brinley, 1959; Heglin, 1956). It may be, however, that like other descriptions and explanations of aging processes, modifiability of behavior is specific to the class of task or situation. A second reason why interpretation of the results of Exp. II need to be qualified is that ". . . slow people are less likely to suffer from interference phenomenon than are fast people . . ." (Brown, 1914, p. 291). This conclusion was reached by Brown when he compared fast and slow performers with respect to a combination speed and error score of sorting a pattern and then sorting a different one. The implication is that if interference is less with slow performers, the relative slowness of the elderly may have hidden a relative inability to alter the task set. It should be noted that with their slower performances, the elderly did not show less interference effects but instead showed nonstatistically greater interference effects.

SUMMARY

The effect of age on speed of card-sorting performance was investigated in relation to difficulty of the task (Exp. I), change in task set (Exp. II), and complexity of the task (Exp. III). In addition, the roles o feducation and consecutive sorting trials were studied in relation to the task variables.

In general, it was found that:

1. Card-sorting speed decreased with increased age, task difficulty, and task complexity.
2. Task complexity tended to make for disproportionate slowing of card-sorting performance in the elderly. Task difficulty had the same effect but only when covaried with complexity. This was most apparent in the task that was both most difficult and most complex.
3. Card-sorting speed decreased also with changed task set. This decrease, however, was independent of age.
4. Rate of improvements in speed of card sorting was greatest with the most difficult and complex task.
5. Age differences were not found in rate of improvement or learning. This was true independent of task difficulty, when the task set was changed, and independent of the complexity of the task.
6. The two levels of education were not significantly different in card-sorting speed.

These results were discussed in the context of age differences in perceptual and cognitive functioning.

REFERENCES

ANDERSON, T. W. *An introduction to multivariate statistical analysis.* New York: Wiley, 1958.

BIRREN, J. E., & BOTWINICK, J. Speed of response as a function of perceptual difficulty and age. *J. Geront.*, 1955, **10**, 433–436.

BOTWINICK, J., BRINLEY, J. F., & ROBBIN, J. S. The interaction effects of perceptual difficulty and stimulus exposure time on age differences in speed and accuracy of response. *Gerontologia*, 1958, **2**, 1–10. (a)

BOTWINICK, J., BRINLEY, J. F., & ROBBIN, J. S. Task alternation time in relation to problem difficulty and age. *J. Geront.*, 1958, **13**, 414–417. (b)

BOTWINICK, J., BRINLEY, J. F., & ROBBIN, J. S. Modulation of speed of response with age. *J. genet. Psychol.*, 1959, **95**, 137–144.

BOTWINICK, J., ROBBIN, J. S., & BRINLEY, J. F. Reorganization of perceptions with age. *J. Geront.*, 1959, **14**, 85–88.

BROWN, W. Habit interference in sorting cards. *Univer. Calif. Public. Psychol.*, 1914, **1**, No. 4, Pp. 269–321.

CLAY, H. M. Changes of performance with age on similar tasks of varying complexity. *Brit. J. Psychol.*, 1954, **45**, 7–13.

CROSSMAN, E. R. F. W., & SZAFRÁN, J. Changes with age in the speed of information-intake and discrimination. *Experientia Supplementum IV.* Symposium on Experimental Gerontology. Basel: Birkhauser, 1956. Pp. 128–135.

HEGLIN, H. J. Problem solving set in different age groups. *J. Geront.*, 1956, **11**, 310–317.

KAY, H. Some experiments on adult learning. In *old age in the modern world.* Edinburgh and London: E. & S. Livingstone, 1955, Pp. 259–267.

WELFORD, A. T. *Ageing and human skill.* London: Oxford Univer. Press, 1958.

(Received January 23, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

ONSET VERSUS TERMINATION OF A STIMULUS AS THE CS IN EYELID CONDITIONING[1]

WESLEY J. HANSCHE[2] AND DAVID A. GRANT

University of Wisconsin

The purpose of the investigation was to determine whether the termination of a visual stimulus had the same effect as the onset in serving as a CS in eyelid conditioning.

The interval between the onset of the CS and the onset of the UCS (CS-UCS interval) has been shown to be critical in conditioning in a number of researches using various trace and delayed CSs and various CRs. Usually a CS-UCS interval of about .5 sec. is optimal (Bernstein, 1934; Kimble, 1947; McAllister: 1953a, 1953b; White & Schlosberg, 1952; Wolfle: 1930, 1932). Indirect conditioning of the pupillary response may be an exception to this general rule (Gerall & Woodward, 1958). The dependency of conditioning upon the interstimulus interval may be used to investigate the relative efficacy as a CS of the termination and onset of a stimulus, because long interstimulus intervals result in very little conditioning or actual extinction (McAllister, 1953b), and unusual latency distributions (Boneau, 1958). In this investigation, therefore, a stimulus of 1.5 sec. duration was used, with the UCS given either .15, .35, .55, or .75 sec. after the onset of the CS and at .15, .35, .55, and .75 after the termination of the CS in eight different groups.[3] If termination of the stimulus is ineffective as a CS, the latter four groups should show less conditioning than the groups with .15, .35, .55, and .75 sec. between onset of CS and onset of UCS. Conversely, if essentially the same function is found relating rate of conditioning to the interval between CS and UCS onset as is found between CS termination and UCS onset, then the argument favors the proposition that onset and termination of a stimulus are equally effective as the CS.

METHOD

Apparatus.—The apparatus was the same as that used in other experiments from the University of Wisconsin laboratories (Grant, Schipper, & Ross, 1952; Prokasy, Grant, & Myers, 1958). The movement of the eyelid was picked up by a microtorque potentiometer,[4] amplified, and differentiated, so as to give an ink recording of the eyelid responses.

The CS consisted of a light in a 10-cm. circular milkglass window which lasted 1.5 sec. with an illuminance of about 160 mL. The UCS was a corneal air puff of .01 to .02 sec. duration and of sufficient strength to elicit a sharp uniform eyelid reflex. The *S*s were run under standard light-adapted conditions in order to minimize sensitization of the

---

[1] This research was supported partly by the Research Committee of the Graduate School of the University of Wisconsin with funds granted by the Wisconsin Alumni Research Foundation and partly by a grant from the National Science Foundation to the junior author.

[2] Now at Tulane University.

[3] Subsequent calibration shows that .04 sec. should be added to all CS-UCS intervals and latencies reported in this experiment.

[4] The use of the microtorque potentiometer which puts virtually no load on the eyelid was originally suggested to the authors by K. W. Spence of the State University of Iowa, but we modified the Iowa arrangements by coupling the shaft of the potentiometer to the eyelid by means of Scotch Tape attached to a length of piano wire which passes into a length of hypodermic tubing, and the tubing is fastened to the potentiometer shaft by means of a hinged coupling.

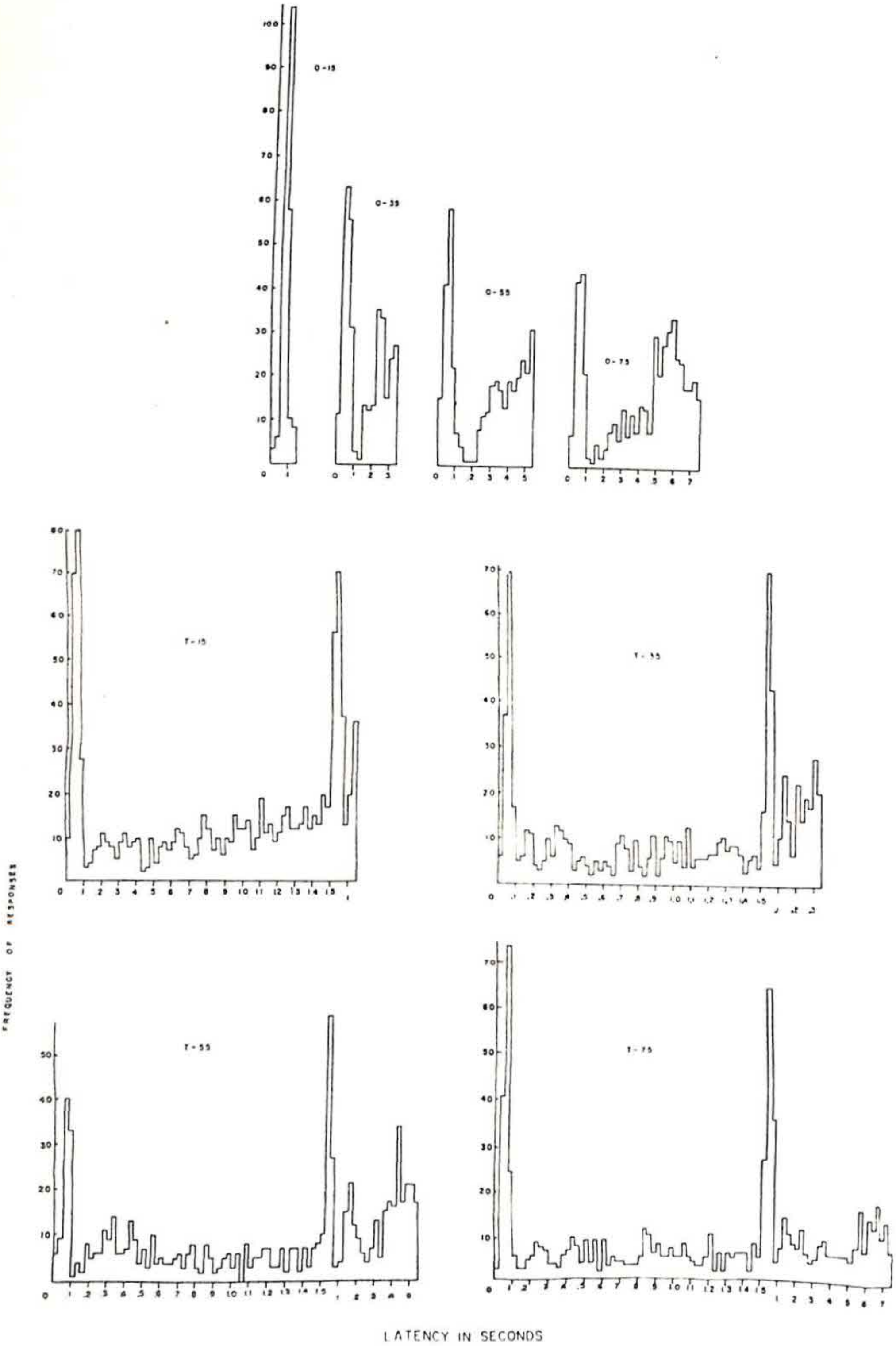


FIG. 1. Latency distributions of all eyelid responses from the four onset groups (above) and the four termination groups below.

beta response (Grant & Norris, 1947). Except in the special groups mentioned below, the corneal air puff was applied to the left eye, and the movements of the right eyelid were recorded.

Procedure.—All *S*s were run for one day only, receiving 40 acquisition trials with 100% reinforcement immediately followed by 10 extinction trials. The intertrial interval varied between 15 and 45 sec. with a mean of about 30 sec.

Experimental design.—The primary experiment was a 2 × 4 orthogonal design, with the four interstimulus intervals of .15, .35, .55, and .75 sec. Four groups (O groups) were run with CS-UCS intervals of .15, .35, .55, and .75 sec. from *onset* of the light stimulus to the onset of the UCS. The second set of four groups (T groups) used the same intervals, but they were measured from the *termination* of the light stimulus. Ten *S*s were randomly assigned to each group by successive replication. Later eight more groups of five *S*s each were run using a very short light stimulus (recorded as .05 sec. but actually shorter due to the latency of the filament of the 60-w. bulb). The CS-UCS interval, measured from actual onset of CS to onset of UCS was .15, .35, .55, and .75 sec., and for each CS-UCS interval one group of *S*s had the UCS delivered to the right eye, and another group had the left UCS delivered to the eye.

Subjects.—The *S*s were 120 men and women students from introductory psychology classes at the University of Wisconsin. They were instructed to maintain a passive attitude toward the experiment and not to influence their natural reactions to the stimuli.

RESULTS

Orderly data were obtained showing that the termination of the stimulus was as effective as its onset in serving as a CS in eyelid conditioning, and, further, that both onset and termination CS-UCS intervals had the same functional relationship to rate of conditioning. As secondary findings, some striking new data on alpha conditioning were obtained, and contralateral conditioning was found to be almost the same as ipsilateral conditioning.

Acquisition.—Latency distributions are presented for all eight groups in the primary experiment in Fig. 1. Here the number of eyelid responses is plotted against latency for all responses during acquisition. In the four onset groups (O groups) the left-hand mode of responses consists of primary reflexes or alpha responses to light and the right-hand mode consists of CRs. For the termination groups (T groups) the left-hand mode consists of alpha responses, the next mode to the right consists of off effect responses (ORs), and the responses further to the right consist of CRs. Occurrence of stimulus termination evidently prevents development of

TABLE 1

ANALYSES OF VARIANCE OF TOTAL FREQUENCY OF ALPHA RESPONSES, OFF RESPONSES, AND CRs DURING ACQUISITION, SUBJECTED TO THE ARCSIN TRANSFORMATION

Source	Alpha Responses		Off Responses		CRs	
	df	*F*	*df*	*F*	*df*	*F*
Onset-Termination (OT)	1	0.10			1	0.13
CS-UCS interval (I)	(3)		(3)		(3)	
Linear	1	5.33*	1	3.38	1	64.75**
Quadratic	1	1.35	1	3.07	1	9.66**
Cubic	1	0.10	1	0.44	1	0.98
OT × I	3	0.46			3	2.38
Error (*MS*)	74	(251.7)	36	(229.6)	72	(220.5)

* $P = .05$.
** $P = .01$.

BUREAU OF EDNL. & PSYL. RESEARCH
1.9.70
LIBRARY
9253

long latency CRs as reported by Boneau (1958). The OR is apparently analogous to the alpha reflex to onset of light. It will be noted that the number of alpha responses declined steadily with increasing CS-UCS interval in the onset groups, showing that the level of alpha conditioning (Hilgard & Biel, 1937; Weber & Wendt, 1942; Grant & Adams, 1944) is related to the CS-UCS interval. This is a new observation. The termination groups show some evidence of this tendency in alpha responding, although the T − .75 group departs from the general trend.

Table 1 presents a summary of the analysis of variance (Grant, 1956) of frequency of responses during acquisition. There is a significant linear trend in the alpha responses as a function of CS-UCS interval. There is no significant interaction between CS-UCS interval and onset versus termination.

Although the alpha responses and ORs seem dependent upon the CS-UCS interval, their primary tendency during the experiment was to adapt out. This is shown in Fig. 2 where both the alpha responses and the ORs decrease to about half the initial frequency during the course of the experiment. There is no evidence of an acquisition trend, so that alpha conditioning is represented primarily by a resistance to adaptation rather than as an actual increase in frequency of responding (cf. Grant & Adams, 1944).

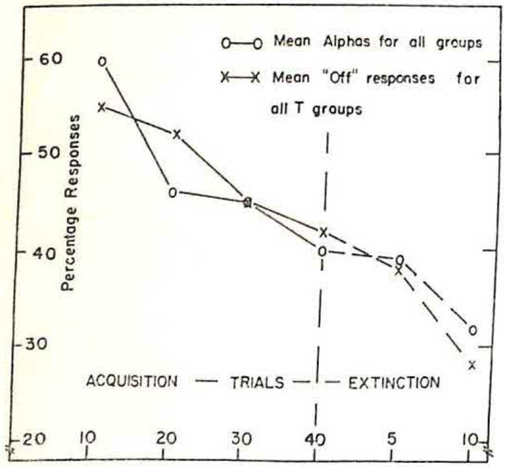


FIG. 2. Percentage frequency of alpha responses and off-responses (ORs) plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction.

The OR follows the termination of the CS at approximately the same latency as that between the onset of the CS and the alpha response. There is no indication that the OR is dependent upon the interval between termination of the CS and onset of the UCS. In Table 1 there are no significant trend components in the frequency of ORs as a function of the CS-UCS interval.

The acquisition and extinction of the CRs in the eight groups are shown in Fig. 3, where percentage CRs is plotted against trials in 10-trial blocks during acquisition and 5-trial blocks during extinction. The usual acquisition trends are shown except in the short CS-UCS interval groups. Progressively more conditioning is found with the higher CS-UCS intervals both with respect to onset and termination, and the linear and quadratic components of this relationship to CS-UCS interval are significant as shown in Table 1. There is no significant difference in the total frequency of CRs between onset and to termination groups.

The CS-UCS interval effect.—Plotting rate of conditioning against the CS-UCS interval during acquisition presents certain technical difficulties. For the short CS-UCS interval the occurrence of the UCS and reflexes to the UCS may hide the CRs so that the level of conditioning cannot be determined directly from the records. In addition, the longer the CS-UCS

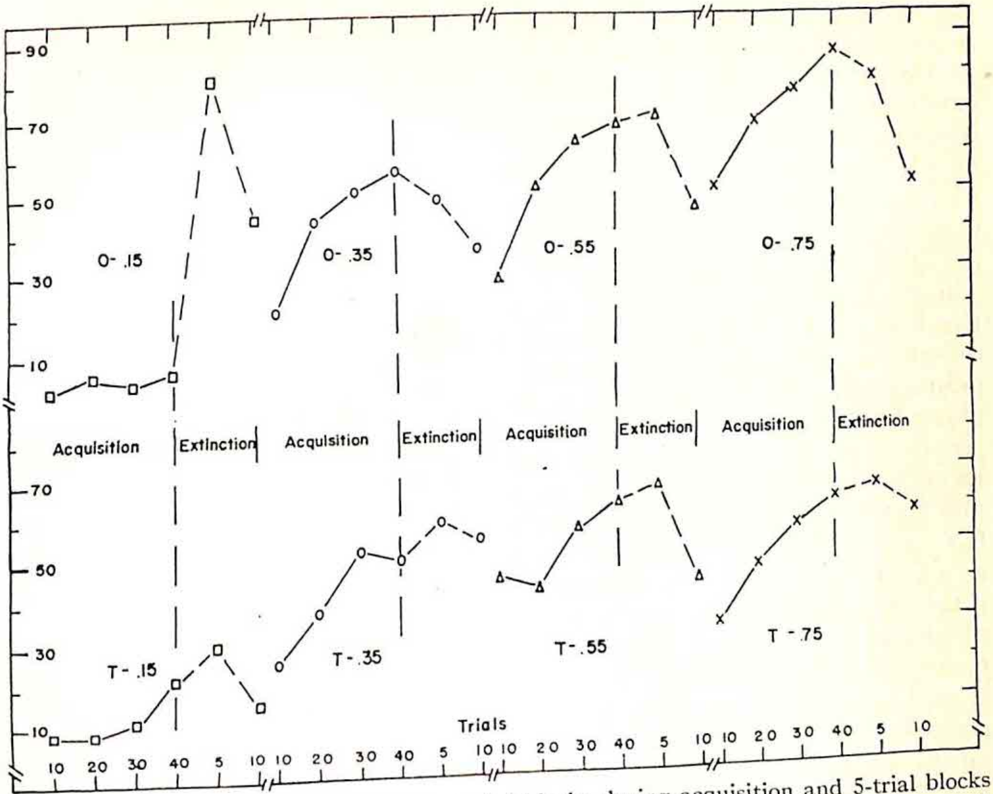


FIG. 3. Percentage CRs plotted in 10-trial blocks during acquisition and 5-trial blocks during acquisition and 5-trial blocks during extinction for the four onset groups (O groups) and the four termination groups (T groups).

interval the greater the opportunity of a random blink occurring during the anticipatory interval before the UCS occurs. Even if increased conditioning did not result in a higher probability of a response occurring during any part of the CS-UCS interval, the longer the CS-UCS interval available for inspection the more responses will be observed. This interval magnitude effect can be handled in various ways. For example, Kimble (1947) inserted test trials on which the UCS was not given in order to get a clear view of all of the CRs. McAllister (1953a) used the first block of acquisition trials to calculate the random rate or operant level of responding during the CS-UCS interval for each *S* and subtracted this amount from the frequency of CRs on the following blocks of trials. In the current experiment a statistical method was employed which may have the advantage of utilizing the data more fully. As an approximation it was assumed that the longer the interval before the UCS the more likely a response was to occur owing simply to the length of the CS-UCS interval and not to the strength of conditioning per se. This implied that over and above general rate of conditioning there would be a linear increase in number of recorded CRs as the CS-UCS interval is increased. Therefore the number of CRs was plotted against the CS-UCS interval and a straight line was fitted by least squares to the observed data points. Assuming the straight line to be primarily interval magnitude effect, the discrepancy between the data points and the straight line was taken as the function relating strength of conditioning to the CS-UCS interval, uncontaminated by the interval magnitude effect. In the upper part of Fig. 4 are shown the data points of the O groups with a fitted line representing the interval magnitude effect. The discrepancy scores are plotted in the lower part of Fig. 4, where the difference score between the data points and the straight lines are plotted against the CS-UCS intervals for the four onset groups, the four termination groups, and the four onset groups with the short duration CS. (In each instance there had been a significant linear and significant quadratic effect so that the significant quadratic and residual variation are displayed in Fig. 4. Casual examination of the figure shows that optimum conditioning was obtained in the neighborhood of .3 to .5 sec. This may be somewhat shorter than the optimum interval that has always been found for the CS-UCS interval function by other investigators in eyelid conditioning using other techniques of correcting for the CS-UCS interval magnitude artifact.)

The most important observation from the standpoint of this experiment is that the termination function in Fig. 4 is essentially the same as the onset function. Furthermore the duration of the CS per se seems to make little difference because the short duration CS produces the same function as the long duration so that trace and delayed functions are essentially identical. This observation is confirmed by the fact that there is no significant interaction between the three groups and the quadratic components of these trends.

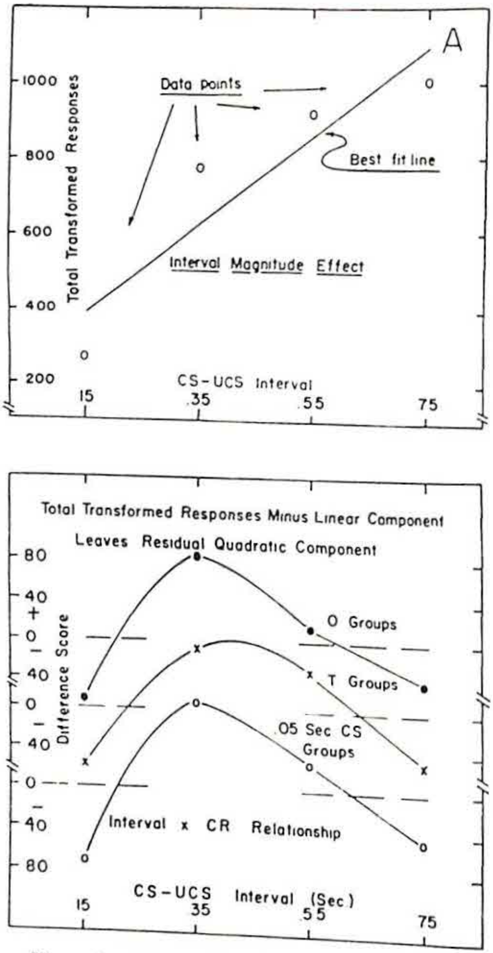


FIG. 4. Determination of the effect of CS-UCS interval on frequency of eyelid CRs. In the upper panel a least squares line representing estimated interval magnitude effect is fitted to the transformed (arcsin) total frequency of CRs in the four onset groups. In the lower panel the discrepancies from the fitted line are plotted along with similar discrepancies from the four termination groups and four short-duration CS groups.

Contralateral vs. ipsilateral conditioning.—Although no analysis will be presented it may be remarked that the acquisition trends and the CS-UCS interval trends were identical for those *S*s whose responses were recorded from the puffed right eye and those whose left eye was puffed and right eye responses were recorded. There was a 5–10% reduction in frequency on the contralateral eyelid. Although the primates generally give a bilateral eyelid response (Hilgard & Marquis, 1936), in the Wisconsin laboratory an occasional *S* is found who gives a greatly reduced contralateral response to puffing the left eye. The results found in this experiment show, however, that contralateral conditioning is about the same as ipsilateral conditioning, except that there is a 5–10% reduction in frequency of CR if the contralateral eyelid movements are recorded.

Extinction.—The distributions of the eyelid response latencies during extinction showed the same modes as during acquisition, although there were fewer responses owing to the shortness of the extinction series and the adaptation of alpha responses and ORs. The frequency of alpha responses, CRs, and ORs during extinction was analyzed as a function of CS-UCS interval in the same way acquisition was analyzed in Table 1. Alpha responses showed the same significant linear decrease with size of CS-UCS interval as was shown in acquisition. Frequency of ORs was not affected by the CS-UCS interval just as in acquisition. In the case of CRs, however, the extinction data showed no significant trend components in contrast to the distinct linear and quadratic trends during acquisition. This was due at least in part to the brevity of the extinction series but also to an "aberrant" result in the extinction of the 0 – .15 group. (See Fig. 3.) This aberrant result was not found with the corresponding group given a short duration CS.

DISCUSSION

The primary finding of the experiment was that the termination of the stimulus may serve almost exactly the same function as the onset of the same stimulus in the occurrence of conditioning. This is reflected in the essential identity of the CS-UCS interval functions for onset and for termination groups. The general implications of this observation are of considerable importance. It has long been known that conditioning is usually optimal when the interval between the onset of the conditioned stimulus and the onset of the unconditioned stimulus is about .5 sec. It is evident from these data that if there are two salient or critical points to the stimulus—in this experiment, the onset and the termination—the conditioning process will be optimal if the UCS is given in the neighborhood of .5 sec. of *either* of these critical features, onset or termination. If the UCS is given .5 sec. after the termination of the CS, the onset of the CS appears to have little or no relevance to the formation of the eyelid CR. In fact a very careful examination of all responses at various latency ranges that occurred during the actual presence of the light stimulus in the termination groups failed to reveal any sign of an acquisition or an extinction tendency among the responses in any latency range. They were, evidently, random blinks. At least they did not show any of the usual tendencies to increase with reinforcement and decrease with extinction.

There is some indication that the short CS-UCS interval, although it does not produce much conditioning in the ordinary sense, does produce a fairly high level of alpha conditioning or reflex sensitization. This was shown during both acquisition and extinction in the onset groups. In spite of the alpha conditioning, the predominant trend of the alpha responses was downward during the experiment, indicating that reflex adaptation had a greater effect than reflex sensitization due to the presence of the corneal puff.

The OR was found to occur with about the same frequency and latency, relative to the termination of the CS, as did the alpha response to the onset of the CS. Beside the almost identical latency, frequency, and adaptation trends of the alphas and ORs there was a correlation of .72 between the frequency of alphas and frequency of ORs for the 40 *S*s for whom both responses were available.

SUMMARY

An experiment with 120 human *S*s was performed to compare the effectiveness of stimulus termination with stimulus onset as the CS in eyelid conditioning.

The principal results were as follows:

1. The effect of the CS-UCS interval on the level of conditioning during acquisition trials was exactly the same whether the onset or the termination of the light served as the CS. Optimal conditioning occurred with a CS-UCS interval of approximately .5 sec. or less.

2. In addition to the usual primary reflex or alpha response to light, an off response (OR) was observed which was closely analogous to the alpha response except that it occurred as a response to termination of the light. The ORs were practically identical with the alpha responses in frequency, latency, and adaptation trend.

3. There was evidence of alpha conditioning that was most marked with the short CS-UCS intervals, but the corresponding trend with ORs was not significant.

It was concluded that the termination of a stimulus is just as critical a feature for conditioning as its onset. If the termination of the stimulus occurs .3–.5 sec. before the UCS, conditioning will take place to termination rather than onset which indicates a certain interesting flexibility in the learning mechanism.

REFERENCES

Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.

Boneau, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

Gerall, A. A., & Woodward, J. K. Conditioning of the human pupillary dilation response as a function of the CS-UCS interval. *J. exp. Psychol.*, 1958, **55**, 501–507.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Grant, D. A., & Adams, J. K. "Alpha" conditioning in the eyelid. *J. exp. Psychol.*, 1944, **34**, 136–141.

Grant, D. A., & Norris, E. B. Eyelid conditioning as influenced by the presence of sensitized Beta-responses. *J. exp. Psychol.*, 1947, **37**, 423–433.

Grant, D. A., Schipper, L. M., & Ross, B. M. The effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

Hilgard, E. R., & Biel, W. C. Reflex sensitization and conditioning of eyelid responses at intervals near simultaneity. *J. gen. Psychol.*, 1937, **16**, 223–234.

Hilgard, E. R., & Marquis, D. G. Conditioned eyelid responses in monkeys, with a comparison of dog, monkey, and man. *Psychol. Monog.*, 1936, **47**(2), (Whole No. 212).

Kimble, A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1–15.

McAllister, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–422. (a)

McAllister, W. R. The effect of eyelid conditioning of shifting the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 423–428. (b)

Prokasy, W. F., Jr., Grant, D. A., & Myers, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55**, 242–246.

Weber, H., & Wendt, G. R. Conditioning of eyelid closure with various conditions of reinforcement. *J. exp. Psychol.*, 1942, **30**, 114–124.

White, C. T., & Schlosberg, H. Degrees of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

Wolfle, H. M. Time factors in conditioning finger withdrawal. *J. gen. Psychol.*, 1930, **4**, 372–379.

Wolfle, H. M. Conditioning as a function of the interval between the conditioned and original stimulus. *J. gen. Psychol.*, 1932, **7**, 80–103.

(Received January 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

THE SOMATIC BACKGROUND OF ROTE LEARNING[1]

R. N. BERRY AND R. C. DAVIS

Indiana University

A previous report (Berry & Davis, 1958) in this series discussed muscle activity in relation to rote learning of nonsense syllables. The present report describes a further experiment on rote learning, with some procedural changes, in which the striate responses of the jaw and arm are re-examined, and heart rate, electrocardiogram, volume pulse, and GSR measures are also evaluated.

As in the previous report, the primary interest was in establishing empirical relations rather than in searching for support for a theoretical position.

METHOD

Subjects and apparatus.—The *S*s were male students, paid volunteers from the summer student population. The cooperation of the *S*s was excellent, and there seems no reason to question their motivation.

Only 16 *S*s are considered in the treatment of the data; there were a number of others who have been excluded because of failure to follow instructions, failure to reach criterion within 25 trials, or malfunction of the apparatus. The 25-trial limitation was set early in experimentation because of the prohibitive cost of operating the recording apparatus for greater numbers of trials.

Each *S* was placed in a chair in the experimental room; the room contained a memory drum and was separated from a recording room by a sound-reducing partition. Bipolar surface electrodes, with saline paste, were fastened over the jaw (masseter) and right forearm (extensor) muscles of *S*. The EKG and the heart rate were obtained from electrodes in the Lead I position, and the volume pulse was obtained with an electric finger plethysmograph. The palmar GSR was recorded from yet another set of leads. The details of the recording apparatus are described in a previous report (Ellson, Davis, Burke, & Saltzman, 1952).

Procedure.—The memory drum was so arranged that *S* was presented with a blank white space in the memory drum for the first 2 sec. This was the interval in which *S* was instructed to make his verbal anticipation of the upcoming item. A light then flashed momentarily in the memory drum aperture to indicate that the anticipation interval was terminated. Two seconds later the item on the drum appeared and remained in view for 2 sec. The 6-sec. cycle was repeated until all of the items had been presented. The next trial was begun without interruption.

The nine items on the memory drum were two-place numbers presented to each *S* in the same order: 35, 29, 26, 32, 39, 27, 24, 38, 31. In the instructions, preceding the learning session, *S*s were told that there was a list of numbers to be memorized and that the numbers might range from 21 through 39. They were also told that there were no duplications within the list. Further instructions were to the effect that the memory drum would continue to rotate until every item in the list was correctly anticipated within the same trial and that *S* was to anticipate each upcoming number, including those on the first trial, even if the anticipations were only guesses. This last instruction was intended to eliminate the difficulty of interpreting the data in those cases where no understandable anticipation was made, a problem which had arisen in the previous experiment.

Data analysis.—The kymograph record contained tracings from all the leads, as well as a tracing indicating the action of the memory drum. Thus the records for an entire learning session could be divided into 1-sec. intervals and the appropriate measurements made for each variable for each 1-sec. interval. The measures are analyzed in terms of the basic 6-sec. cycle of the memory drum, covering the anticipation, delay, and information intervals. Further analyses are made in terms of correct and incorrect anticipations and in relation to speed of learning. The latter is defined by the number of trials to a criterion of one correct trial, and four learning speeds are differentiated: 6–10 trials,

---

[1] This research was carried out under Contract Nonr 908(03). R. C. Davis is the principal investigator under this contract. The senior author is indebted to the Graduate School of Indiana University for financial assistance in carrying out this study.

11–15 trials, 16–20 trials, and 21–25 trials. There are 6, 5, 2, and 3 *S*s in these respective groups, and they are identified hereafter as Groups I, II, III, and IV.

RESULTS

The MAP from the masseter.—The recording device integrated the output voltage of the masseter for .1-sec. intervals. Thus for any 1 sec. of the kymograph record, there were 10 values available. In all cases, only the largest of these 10 values has been considered in the data treatment. This practice is in line with previous studies and assumes that it is the maximum response in which the primary interest lies.

Six MAP values (representing each second of the 6-sec. memory drum cycle) for the correct anticipations and six values for the incorrect anticipations were obtained for each *S*. This was done by first obtaining medians across serial positions for each trial and then medians across trials.

The *S*s were grouped according to learning speeds as indicated above, and the mean values, second by second, for each group for correct and incorrect responses are presented in Fig. 1. A variety of hypotheses might be tested, but most importantly those that were supported in the earlier study.

It is quite apparent, in the present experiment, that there is a tendency for the over-all activity to decrease as learning speed decreases; the poorest 5 *S*s (Groups III and IV) show very little activity throughout. The same trend, with a reversal for the very poorest *S*s, existed in the earlier data, and a significant relation between speed of learning and an over-all measure of jaw-forehead activity was also found. In the present experiment, only jaw data are available. However, for the over-all mean for each *S*, the Kruskal-Wallis test yields $.06 > P > .05$. This is very close to the value obtained in the previous experiment, and, correcting for the small *N*s in some of the groups, is significant at the .05 level. That there are no high voltages for the slowest learners in this study, as opposed to the earlier study, might be attributed to the truncated distribution produced by considering only those *S*s who reached criterion within 25 trials.

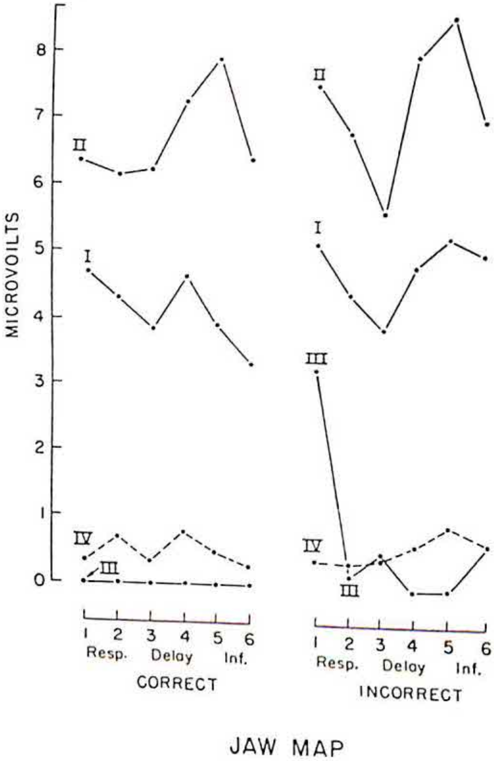


FIG. 1. The mean MAP values for each second of the 6-sec. memory drum cycle for correct and for incorrect anticipations. The roman numerals identify the learning groups, I being the fastest and IV the slowest.

It is possible that the significant correlation between learning speed and general level of tension in the masseter is a function of temporal changes over trials. However, this would seem an unlikely explanation since the over-all tension levels for the correct and incorrect responses appear to be very similar within learning groups. If temporal factors were significant, the incorrect responses, tending to come earlier in the series, would show a high over-all level since the rapid learners showed a higher over-all level.

In the earlier study there was a significant nonmonotonic relation between speed of learning and the difference between the changes from response to information interval for correct and for incorrect responses. In the present experiment this trend is confirmed, but it is not significant at the .05 level. The small difference in results between the two experiments might easily be attributed to procedural differences, making a clear-cut comparison difficult.

The MAP from the arm extensor.—The data can be discussed very briefly. As in the previous experiment in which the arm MAP was recorded during rote learning, no systematic relation between the muscle action measure and any learning measure could be demonstrated.

The electrocardiogram.—The measure of the EKG was the height of the line from the Q to the R points, and this height was converted to a percentage of the mean rest height. The mean rest height was the mean of the values for seven consecutive heart beats occurring about 30 sec. before the learning session was begun. Because the measure is relative, it is not appropriate to make any interpretation which assumes absolute values.

Figure 2 shows the mean second by second values across all serial positions and all trials for each of four learning groups, with a further breakdown into correct and incorrect responses. There is a tendency for the information interval to show large values, and this is quite striking in the case of the incorrect responses. Since these trends occur for all learning groups, the data for all *S*s are combined and presented in Fig. 3. The most striking feature in this plot is the difference between the correct and the incorrect response magnitudes in the information interval. If we apply a simple binomial sign test, the mean of Sec. 5 and 6 values is significantly larger ($P < .01$) for the incorrect responses than for the correct responses. (There is a tendency, although not significant at the .05 level, for the size of this difference to be larger for the rapid learners.) A more rigorous argument concerning the significant finding is made by a further manipulation of the data. For each *S* the mean value of Sec. 5 was subtracted from the mean value of Sec. 4 for the correct responses and for the incorrect responses. The binomial sign test shows a significant difference ($P < .05$) between the difference measures for the correct and for the incorrect responses.

From what is known of the relation between respiratory cycle and the

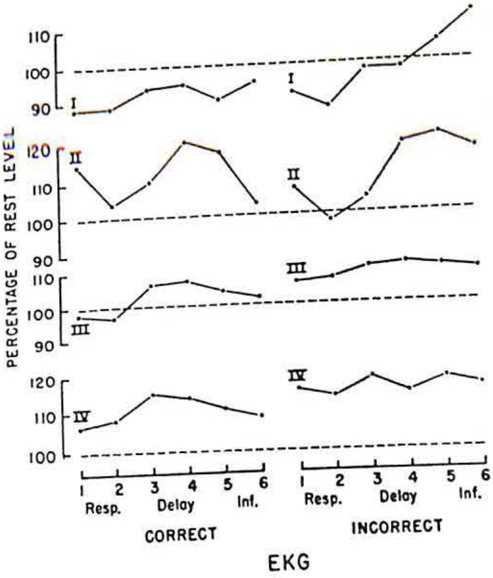


FIG. 2. The electrocardiogram values for each of the four learning groups.

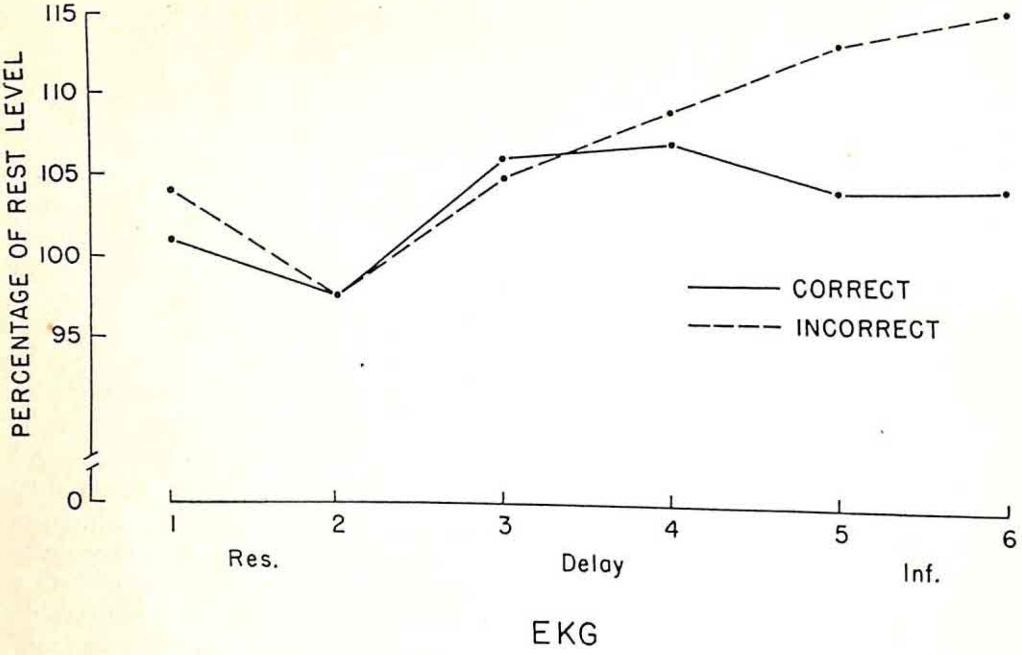


FIG. 3. The data of Fig. 2 combined for all learning groups.

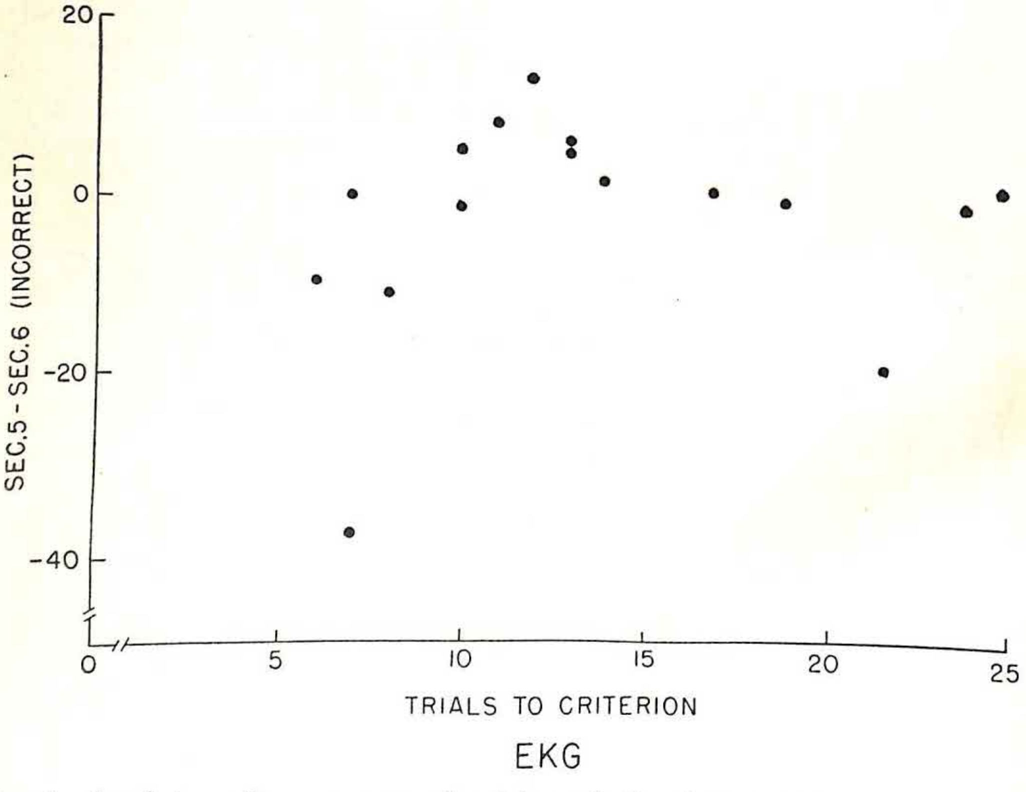


FIG. 4. An electrocardiogram measure (involving only Sec. 5 and 6 of the memory drum cycle) related to speed of learning.

electrocardiogram, it is safe to assume that the larger measures represent an expiratory movement in the breathing pattern. The data would sensibly be interpreted as corrective responses in the case of an incorrect anticipation, the corrective responses otherwise not being required. It may be noted in Fig. 3 that the differences between the values of the rights and wrongs in the response interval are negligible. This gives further emphasis to the importance of the information interval in the learning pattern.

Since the information interval for the incorrect responses appeared to be of interest, one further analysis was made utilizing the Sec. 5 and 6 values. The data for the analysis are summarized in Fig. 4. For the incorrect responses, the mean value of Sec. 6 is subtracted from the mean value of Sec. 5 for each *S* and plotted against the number of trials to criterion. It is apparent that the relation is curvilinear; the Kruskal-Wallis test shows that the relation is significant at the .05 level. Since there is no such systematic relation for the correct anticipations, the nature of somatic responses in the information interval following an incorrect response appears to be critical.

The volume pulse.—The magnitude of the volume pulse was measured for the first pulse in each 1-sec. interval and converted to a percentage of rest level. The latter was a mean magnitude for six consecutive seconds in a period about 30 sec. before the memory drum was started.

The mean data for the correct and incorrect responses are presented in Fig. 5. Here again, *S*s are divided into four learning groups, the same ones as used above. Although there

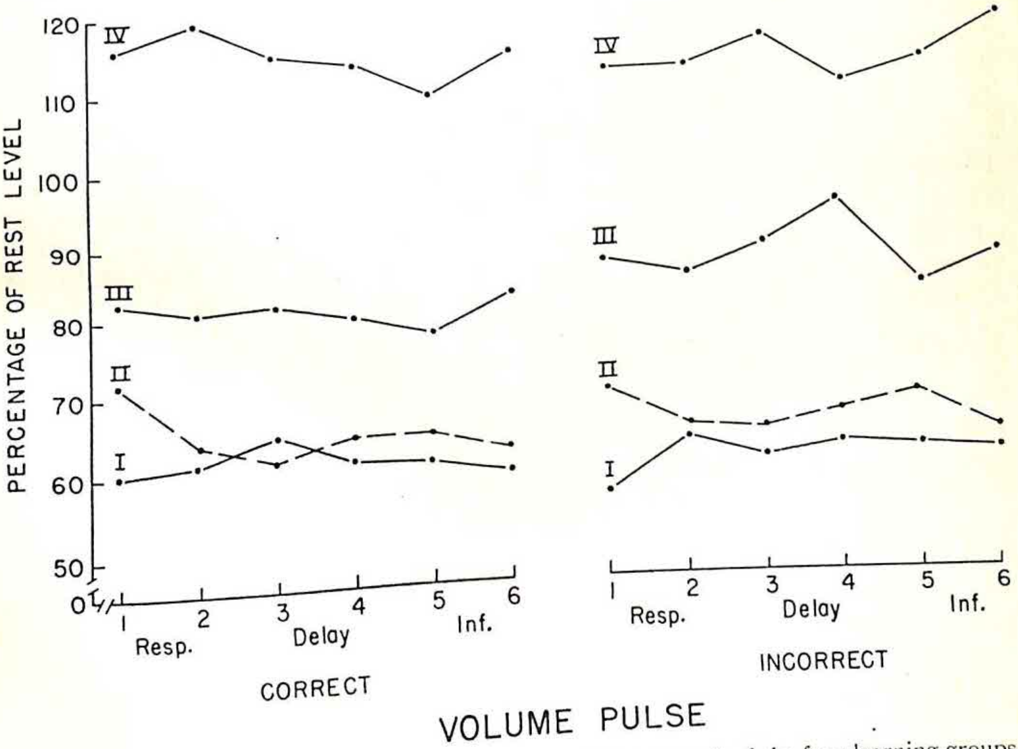


FIG. 5. The mean volume pulse (percentage of rest level) for each of the four learning groups.

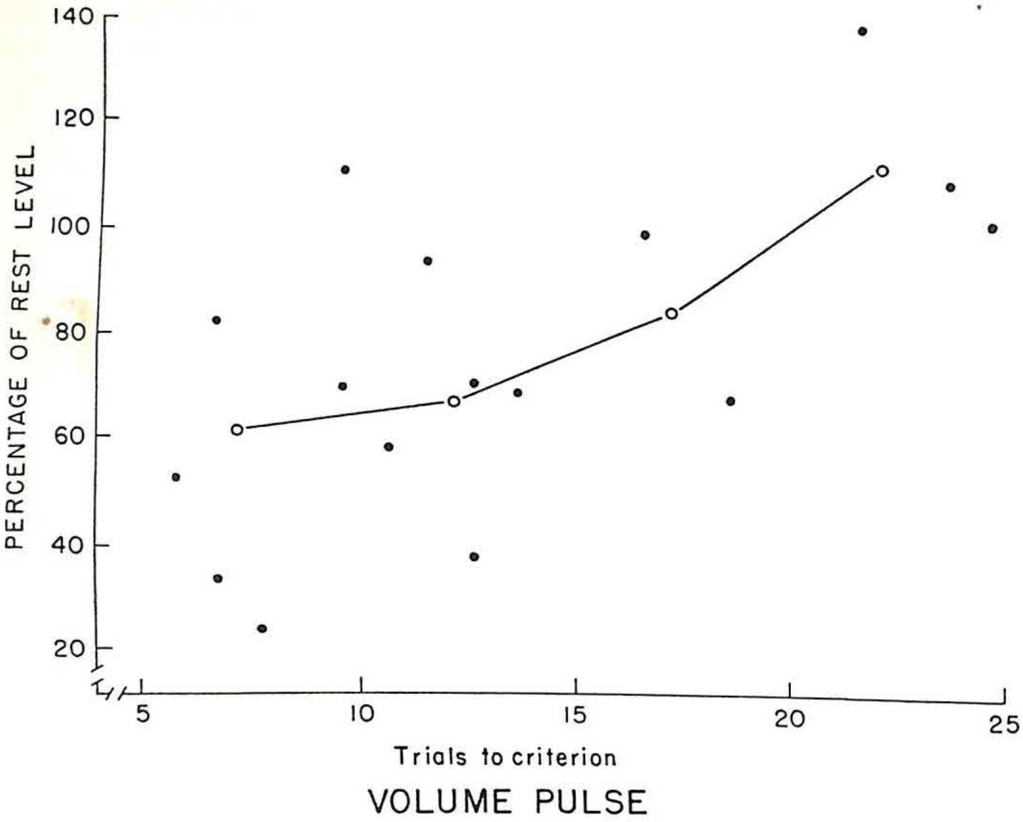


FIG. 6. The over-all mean volume pulse for each *S* related to speed of learning.

are no apparent trends across the 6-sec. cycles for any of the four groups, there is a trend from the best to the poorest learners. Since the trends for correct and incorrect response values are similar, and since there is no systematic relation between position of the memory drum cycle and volume pulse, all readings for each *S* were averaged to give a single value. These final values are presented in Fig. 6; it may be noted that the majority of values represent decreases, rather than increases, from the pre-learning level of the volume pulse. This is particularly true of the better learners. The hollow circles are the means for the four learning groups, and it is these points which are connected to show the over-all trend. The rank-order correlation between the magnitude of the volume pulse and learning speed is .64 ($P < .01$). It can reasonably be asked whether the correlation is influenced by a simple trial to trial pattern of volume pulse change. To answer this, mean values of the first three trials were obtained for each *S* and then averaged across *S*s to get values for each of the four learning groups. The same procedure was used for the middle three and last three trials. The trends (Fig. 7) show no systematic slope changes, thus indicating that trial to trial changes in volume pulse during learning are not the basis for the correlation between volume pulse and speed of learning.

Heart rate.—The heart rate measures were in terms of time between beats, which has a reciprocal relation to the usual measure of beats per

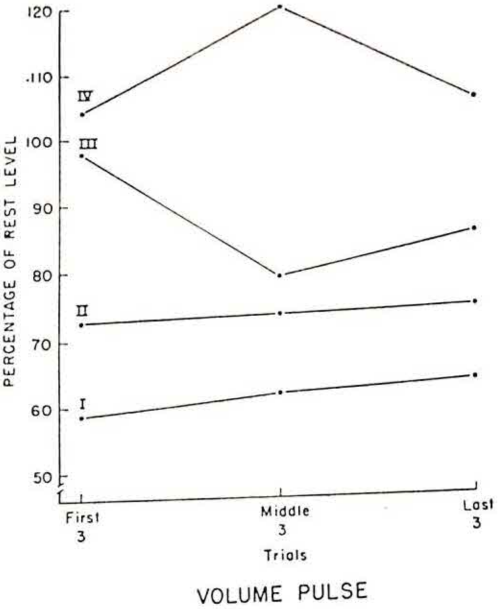


FIG. 7. A mean volume pulse measure for each learning group related to trial position.

minute. Thus the larger values reflect slower heart rates.

The mean data for the four learning groups for both right and wrong responses are presented in Fig. 8. Grouping all *S*s, there is a significant trend for the heart rate to decrease from Sec. 1 to Sec. 5 of the cycle, for both correct and incorrect responses. However, the magnitude of the decrease is not significantly related to speed of learning for either response category, nor are there significant differences in the trends for correct and incorrect responses. Lastly, using a mean heart rate involving all the values for each *S*, no significant correlation between these means and speed of learning could be obtained.

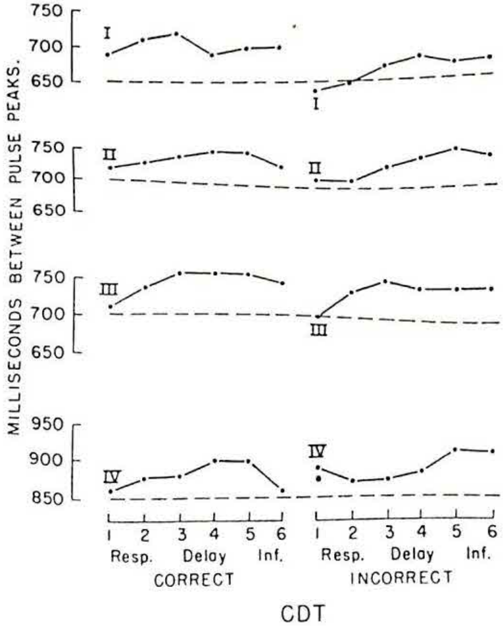


FIG. 8. The heart rate measure for each of the four learning groups.

Heart rate changes are probably a secondary effect of an increased EKG. Stronger heart action would slow pulse rate through the action of the carotid sinus reflex. As a secondary effect, the heart rate changes would have low or negligible correlation with other changes.

GSR.—The GSR records were measured in two ways. The absolute levels of resistance were determined for the moment at which the memory drum was started and for the moment at which it was stopped. There is a nonsignificant trend for the rapid learners to exhibit higher skin resistances. The second measure of the GSR was the number of times that a decrease in skin resistance occurred in the first trial, in the trial half way through the learning session, and in the last trial. Again, no significant or systematic relations among the learning speeds or among the trial positions are evident. There were very large differences among *S*s for both GSR measures, differences which would make the negative findings quite tentative.

DISCUSSION

The physiological adjustments to the task situation as well as the moment to moment changes have been shown to relate to measures of learning performance. As for the adjustment to the task situation, speed of learning correlates with the magnitudes of the jaw MAP and of the volume pulse. Generally, it seems that the most rapid learners generate a high striate activity in the head region.

The moment to moment changes associated with speed of learning are reflected in the jaw MAP. There is a tendency for the MAP pattern differences between the correct and incorrect anticipations to vary systematically with speed of learning. The same tendency occurs in the EKG data. And the EKG exhibits a significant correlation between temporal patterns for the incorrect responses and speed of learning.

As an over-all generalization, it is contended that the rate of verbal learning is associated with increased physiological activity in the head region and further that this activity is most important in the interval when a correcting response is made following an incorrect anticipation.

SUMMARY

Sixteen *S*s learned under instructions to verbally anticipate every item in a list of nine two-place numbers; the learning criterion was one correct trial. The memory drum presented a basic 6-sec. cycle for each number: a 2-sec. anticipation interval, a 2-sec. delay interval, and a 2-sec. information interval. There was no intertrial interval. The following variables were recorded on kymograph tape: jaw MAP, arm MAP, heart rate, EKG, volume pulse, GSR.

The data support the following statements:

1. In the over-all adjustment to the task situation, speed of learning has a significant positive correlation with magnitude of jaw MAP and with volume pulse.

2. There is a tendency for the difference in MAP (jaw) patterns for correct and incorrect responses to vary systematically with individual differences in learning speed. The same tendency occurs in the EKG data.

3. The temporal pattern of EKG magnitudes during the information interval of incorrect anticipations shows a significant correlation with speed of learning.

4. It is concluded that the rate of verbal learning is associated with increased physiological activity in the head region and further that this activity is most critical in the interval when a correcting response is made following an incorrect anticipation.

REFERENCES

BERRY, R. N., & DAVIS, R. C. Muscle responses and their relation to rote learning. *J. exp. Psychol.*, 1958, **55**, 188–194.

ELLSON, D. G., DAVIS, R. C., BURKE, C. J., & SALTZMAN, I. J. *A report of research on detection of deception.* Contract N6onr-18011, Indiana University, 1952.

(Received January 31, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

"ASSOCIATION" OF NONSENSE SYLLABLES FOLLOWING VARIED LEARNING CONDITIONS[1]

BONNIE WEBB CAMP[2]

Indiana University

This study was designed to determine whether some of the assumptions concerning the development of word associations and of verbal behavior in general could be investigated directly in nonsense syllable learning experiments which use the association method as a test of learning. The association method itself consists of presenting *S* with a stimulus, usually a word or syllable, with instructions to respond with the first other word or syllable that occurs to him. Traditionally, these responses are called "associations," and in this paper, the term associations will be used only in reference to responses obtained by this procedure.

Even in this narrow sense, however, associations, particularly word associations, have been the subject of considerable interest in both experimental and clinical settings. In spite of the obviously learned nature of word associations, however, the association method has not been used as a test of learning outside of studies of intraserial learning phenomena (McGeoch & Irion, 1952). Although more recent studies (Jenkins & Russell, 1952), have departed from the classical studies of classification and "emotional disturbance" (Bell, 1948; Woodworth, 1938), attempts to observe directly, through experimental manipulation, the conditions of learning which are functionally related to the production of associations have been notably lacking. It would appear that investigating the acquisition of associations in an experimental setting might help to increase our understanding of the role of antecedent events in the production of word associations and of verbal behavior in general. The present study employing nonsense syllables represents a step in this direction.

Trials with paired-associate nonsense syllables provides, in a learning task, the basic elements which are tested by the association method. Presentation of syllable pairs, however, does not in itself eliminate the possibility that differences in associations could be attributed to differences in *S*'s exposure to the syllables in an experimental situation and, consequently, his familiarity with them. If we define familiarity with the syllables as the frequency of occurrence of the syllables as responses, whether alone or in pairs, familiarity with either member of a pair of syllables becomes a potentially important factor which could have an independent effect on results of an association test.

The following study, therefore, was designed to determine the effect of variation in the number of paired presentations of syllables and variation in familiarity with either member of a pair of syllables on the syllables produced in an association test.

---

[1] This study is based on a dissertation presented to the Graduate School of Indiana University in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express appreciation to A. M. Buchwald, committee chairman and to D. C. Beier, I. J. Saltzman, and H. E. Driver.

[2] Now at University of Colorado School of Medicine.

METHOD

The general plan of this study involved visual presentation of nonsense syllables in varying frequencies of occurrence, singly and in pairs. The *S*s were simply asked to copy the syllables on separate sheets of paper as they were exposed. This insured that each *S* used each syllable, or pair of syllables, as a response the required number of times and kept his method of responding consistent with the method of presentation. After *S*s were presented with single syllables in a random sequence, followed by pairs of syllables in a random sequence, an association test of learning was introduced.

Apparatus.—Twelve nonsense syllables with 40% association value were chosen from Glaze's (1928) list of syllables. Two lists of six syllables each were selected at random within the limits of criteria insuring the least possible amount of repetition of vowel and consonant sounds. Pairs of syllables consisted of one syllable from each list. Syllables in one list were designated stimulus syllables (S syllables). Those in the other list were designated response syllables (R syllables). The syllable pairs were referred to as S-R.

The selected syllables were printed in the center of $1\frac{1}{2} \times 2\frac{1}{2}$-in. cards. The S cards were those bearing S syllables, and R cards those bearing R syllables. On each S-R card, an S syllable and an R syllable were printed side by side with the S syllable always appearing on the left. The same R syllable was always paired with a given S syllable and thus gave a total of six different S-R cards.

The syllables were then arranged in three different types of series. The first type consisted of nine sets of both S and R cards. Each set contained either 2, 6, or 10 occurrences of each S syllable, and either 2, 6, or 10 occurrences of each R syllable arranged in a random order. All *S*s receiving the same combination of S and R occurrence were presented with the same set of cards in the same order.

The second type of series consisted entirely of the six S-R cards. These cards were ordered randomly in sets of six cards so that each S-R pair occurred once in each set. Cards were arranged in a total of 10 such sets. By presenting either none of these sets, the first 6, or all 10 of these sets, it was possible to present different groups with 0, 6, or 10 exposures of each S-R pair. The nine sets containing both S cards and R cards and the series of S-R cards were combined in such a way that each group of five *S*s received 2, 6, or 10 exposures of each S syllable, 2, 6, or 10 exposures of each R syllable and 0, 6, or 10 exposures of each S-R pair. This resulted in a total of 27 possible combinations of each condition.

The third type of series consisted of one each of the six different S cards arranged in a random order. These cards were presented on test trials to all *S*s.

Subjects.—The *S*s were 135 undergraduates enrolled in introductory psychology classes at Indiana University. They were divided at random into 27 groups of five each so that each of the possible combinations described above was administered to five *S*s.

Procedure.—The *S*s were seen individually in a small room ordinarily used for administering psychological tests. They were seated comfortably at a small table across from *E*. A lower table was placed beneath this table and to one side of *E*. This allowed *E* to conceal from *S* all cards except the one being presented. A stack of $2\frac{3}{4} \times 2\frac{1}{4}$-in. paper and a pencil were placed on the table in front of *S*. The room was otherwise bare.

Each *S* was read the following instructions: "I have a series of cards here with nonsense syllables on them. Do you know what a nonsense syllable is?" (If *S* indicated that he did not know, an explanation was given.) "I am going to show these to you one at a time. Write each one on the top piece of paper, turn the paper over immediately and do not look at it again. All you have to do is write each one down." The *S*s were then shown one of the sets of cards containing both S and R cards. The *E* presented the cards manually, one at a time. Each card was removed and replaced by another when *S* had finished copying the syllable.

After completion of the first series, paper on which *S* had written was removed. Those *S*s receiving a series of S-R pairs were read the following instructions: "I have another series of cards but this time there are two nonsense syllables on each card. Write both of them on the top piece of paper, turn the paper over immediately and do not look at it. Again all you have to do is write them down."

Following this, all paper was removed from the table and a pad containing 12 sheets given to *S*. For those *S*s who received no S-R pairs, this procedure followed completion of the first series. These instructions were read to all *S*s: "I am going to show you some of these same syllables again. This time, on the top sheet of paper, write the first other nonsense syllable that you think of, fold the paper over immediately and do not look at it again. All you have to do is write down the first nonsense syllable that comes to mind or that occurs to you, as long as it is not the same as the one on the card you are looking at; and as long as it is one of the syllables you have been writing down."

Results and Discussion

A total of 810 associations was obtained from the group of 135 *S*s. Of these, 290 were S syllables, 403 were R syllables and 117 were extraneous syllables. Of the total number of R syllables, 175 were correct reproductions of the paired associate of the S syllable being presented. Although these frequencies cannot be subjected to a statistical test, the probability of obtaining this large a number of R, S, or correct syllables by chance, is, at most, very small. Furthermore, no *S* gave only extraneous syllables, and many gave none. Chi squares revealed that these extraneous syllables were not distributed in a manner significantly different from chance for any of the three main treatment variables and consequently these responses will not be considered further.

Each *S* was, therefore, given a score corresponding to each of the following: (*a*) number of R associations, (*b*) number of S associations, and (*c*) number of correct associations. Table 1 presents the means of the R and S scores for groups classified on the basis of R familiarity and frequency of S-R presentation.

With one exception, Table 1 shows a progressive increase in R scores and a decrease in S scores as a result of increasing R familiarity and increasing S-R occurrence. The exception to these trends occurs in the six S-R group least familiar with R syllables. This group produced more R associations and fewer S associations than expected from surrounding groups. The S scores are also observed to be higher than R scores in the zero S-R group, whereas R scores are higher than S scores in the groups exposed to S-R pairs. The exception to these trends occurs in the zero S-R group most familiar with R syllables where R scores are slightly higher

TABLE 1

Means and *SD*s of R and S Scores for R Familiarity and Frequency of Paired Presentation

Scores	R Familiarity							
	2		6		10		Total	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
R Scores								
S-R Trials:								
0	1.1	1.3	2.3	1.4	2.9	1.6	2.1	1.6
6	3.3	2.0	2.8	1.5	3.5	1.4	3.2	1.6
10	2.9	1.5	3.7	1.2	4.4	1.2	3.7	1.4
Total	2.4	1.9	2.9	1.5	3.6	1.5	3.0	1.7
S Scores								
S-R Trials:								
0	4.5	1.5	2.7	1.8	2.4	1.6	3.2	1.9
6	1.3	1.3	2.0	1.2	1.7	1.7	1.7	1.4
10	1.5	1.6	1.9	1.5	1.1	1.1	1.5	1.3
Total	2.4	2.6	2.2	1.5	1.7	1.5	2.1	1.7

TABLE 2

ANALYSIS OF VARIANCE OF R AND S SCORES ON ASSOCIATION TEST TRIALS

Source of Variation	*df*	*MS*	*F*
Between *S*s			
Stimulus familiarity (A)	2	1.233	.74
Response familiarity (B)	2	2.211	1.33
S-R frequency (C)	2	1.300	.78
A × B	4	1.477	.90
A × C	4	.611	.37
B × C	4	1.516	.91
A × B × C	8	.336	.20
Error (Between *S*s)	108	1.657	
Total Between *S*s	134		
Within *S*s			
S and R scores (D)	1	47.293	23.24***
A × D	2	19.471	9.57***
B × D	2	20.336	9.99***
C × D	2	66.382	32.62***
A × B × D	4	3.404	1.67
A × C × D	4	5.614	2.76*
B × C × D	4	11.787	5.79***
A × B × C × D	8	6.251	3.07**
Error (within *S*s)	108	2.035	
Total within *S*s	135		
Grand total	269		

* $P = .05$.
** $P = .005$.
*** $P = .001$.

than S scores. With the exception of the zero S-R group at the lower two levels of R familiarity, where *S* scores are higher than R scores, R scores are higher than S scores throughout Table 1.

Results of an analysis of variance performed on the R and S scores is summarized in Table 2.[3] The absence of significant variation arising from Between-*S*s sources of variation suggests that *S*s not only learned the syllables but produced learned syllables consistently at all levels of the three main treatment variables. Examination of Within-*S*s variation, however, reveals significant differences between the R and S associations and all but one of the interactions between this variable and the three main treatment variables.

The over-all trend for *S*s to give more R than S associations is reflected in the difference between R and S scores significant well beyond the .001 level. This suggests that *S*s distinguished between the two classes of R and S syllables as a result of exposure to S-R pairs. Since two-thirds of the *S*s were so exposed while one-third was not, a larger total number of R associations would be expected on this basis alone. This is reflected in the C × D interaction which is significant at well beyond the .001 level. In addition, the B × D interaction, also significant at the .001 level, reflects the trend toward higher R scores with increasing R familiarity. The B × C × D interaction, also significant at the .001 level, appears to result from the unusually high S scores observed at the lower level of R familiarity and zero S-R occurrence and the unusually high R scores observed at the highest level of R familiarity and S-R occurrence.

An increase in S scores with increasing S familiarity accompanied by high R scores at the lowest level of S familiarity is reflected in the A × D interaction, significant at the .001 level. The A × C × D interaction, significant at the .05 level, seems to reflect primarily the occurrence of unusually high S scores at the highest level of S familiarity in the absence of paired presentation and the unusually high R scores at the lowest level of S familiarity when pairing is present.

Since neither the A × B × C nor the A × B × D interaction approaches significance, the A × B × C × D interaction, significant at the .005 level, seems to reflect primarily the relative dominance of the effects of pairing over those associated with familiarity when these variables have opposite effects on the production of R and S scores. In spite of the facilitating effect on S scores associated with increasing S familiarity, S scores are higher than R scores only in those groups receiving no exposure to S-R pairs at the two lower levels of R familiarity. The facilitating effect of both R familiarity and S-R presentation on the production of R associations is sufficiently large to result in higher R scores everywhere else.

Since only S familiarity shows a facilitating effect on the production of S scores and all levels of S and R familiarity are present in the groups receiving no S-R pairs, the predominance of S scores even at the zero S-R level cannot be attributed to greater familiarity with the S syllables alone. Apparently, some other factor, operating with little strength once exposure to pairs is introduced, exerts strong influence on associations in the absence of pairing. A reasonable accounting of this is provided by reference to a process which Skinner (1957) terms "formal prompting." Thus, although *S*s in the zero S-R group had no reason to divide the syllables into R and S classes, presentation of the S syllables as stimuli in the association test could function to prompt these syllables on subsequent trials. On the second test trial, for example, the probability of obtaining the first trial's S syllable as an association would be raised. On the third test trial, the probability of obtaining the second trial's S syllable would be raised. On the last test trial, all but one of the S syllables would have been so prompted. In *S*s receiving no S-R pairs, this could account for the proportionately greater strength of the facilitating effects of S familiarity when R familiarity is low and the balancing of effects of R familiarity when R familiarity is high.

---

[3] The author wishes to express appreciation to W. L. Sawrey for his statistical design for this analysis and to Max Zwanziger for completing the computational procedures.

Table 3 shows the distribution of correct associations for different levels of S-R occurrence. No *S* in the zero S-R group gave more than one correct association and more than two-thirds of *S*s in this group gave none. In groups exposed to syllable pairs, the number of correct associations increases as frequency of S-R presentation increases.

TABLE 3

DISTRIBUTION OF CORRECT ASSOCIATIONS FOR S-R FREQUENCIES

S-R Frequency	No. Correct Associations		
	0	1	2
0	34	11	0
6	16	9	20
10	10	10	25

TABLE 4

MEDIAN TEST OF CORRECT ASSOCIATIONS FOR THE THREE MAIN TREATMENTS

Source of Effect	*df*	χ^2
Stimulus familiarity (A)	2	1.260
Response familiarity (B)	2	.720
S-R frequency (C)	2	28.080*
A × B	4	5.007
A × C	4	1.999
B × C	4	.846
A × B × C	8	8.444
Total	26	46.356

* $P = .001$.

Since homogeneity of variance could not be established for correct scores, the nonparametric test of the medians described by Wilson (1956) for use in analyzing a three-factor design with replication was performed on these scores. A summary of this analysis is presented in Table 4. The trends noted in Table 3 are reflected in the chi square for the S-R frequency variable, which is significant well beyond the .001 level. Neither S familiarity, R familiarity nor any of the interaction effects approaches significance.

Although the trends associated with the S-R variable are in the expected direction, the absence of significant effects arising from variation in R or S familiarity is particularly interesting in view of the findings with respect to the two general classes of R and S syllables. The results of the analysis of correct scores is interpreted as providing further evidence that the effects of exposure to pairs of syllables takes precedence over both the effects of familiarity alone and any prompting effects inherent in the association method.

The importance of this study is seen as resting primarily in the presentation of a method for studying the acquisition of associations under experimentally controlled conditions. The over-all results clearly indicate that variations in responses obtained by the association method reflect, systematically, variations in conditions of incidental, paired-associate learning of nonsense syllables. It does, therefore, seem reasonable to consider further the potential of this method for testing hypotheses derived from word association studies. Extension of the method to any study in which more than one R syllable is paired with a given S syllable, or any study in which more than one S syllable is paired with a given R syllable will, however, necessitate the confounding of either S or R familiarity with frequency of paired presentation. To provide a guide for designing and interpreting such extensions, the present study focused on exploring the systematic effect either R or S familiarity might be expected to produce independently of pairing. Findings that S familiarity is significantly related only to the frequency of S syllable associations and that both increased R familiarity and pairing were related to decreased frequency of S associations, as well as increased frequency of R associations, provides a basis for deciding how to introduce controls for familiarity in future use of the method. For example, if one focuses on studying variation in responses which can be unequivocally classified as R associations or "not S associations," it should be unnecessary to introduce independent control over S familiarity; but, provision for independent control over R familiarity should be included. If, on the other hand, more than one S syllable is paired with a given R syllable, to study the variation in S associations when the R syllable is used as the stimulus in the association test, independent control over both R and S familiarity should be introduced.

SUMMARY

This study presents a method for investigating, experimentally, the formation of associations through use of the association method as a test of incidental, paired-associate learning of nonsense syllables. Focus was placed on determining the relationships between variations in stimulus syllable familiarity, response syllable familiarity and frequency of pairing, and variations in associations. When associations could be viewed as a function of familiarity alone, syllables were given more frequently as associations as familiarity increased. Introduction of pairing tended to reinforce the tendency to give response syllables whether response syllables were more familiar or not and to decrease the tendency to give stimulus syllables whether these were more familiar or not. Pairing, further, was the only variable which related significantly to the production of correct associations.

REFERENCES

BELL, J. E. *Projective techniques.* New York: Longmans, Green, 1948.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–822.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

SKINNER, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received February 11, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

PREDICTIONS OF SEQUENCES OF SUCCESSES AND FAILURES

CLINTON B. DE SOTO, EDMUND B. COLEMAN, AND PETER L. PUTNAM

Johns Hopkins University

In most research on levels of aspiration, *S* selects his level of aspiration from an achievement scale having many discriminable values (Lewin, Dembo, Festinger, & Sears, 1944; Simon, Shaw, & Gilchrist, 1954). The games he plays or the tests he takes have 10, 100, or more, possible scores. In many settings outside the laboratory, however, people are faced with an achievement scale having only two discriminable or acknowledged values, success and failure. A salesman either lands the customer or he doesn't. And in these cases, too, people set levels of aspiration in the sense that they expect either success of failure. As a salesman attempts to make a sale, he may expect either to succeed or to fail at it.

If people repeatedly face the same or similar tasks, succeeding a certain proportion of the time and failing the rest of the time, they will expect success a certain proportion of the time and failure the rest of the time. How is the proportion of trials on which they expect success related to the proportion of trials on which they actually succeed? The present article reports an attempt to answer this question experimentally.

It is proposed that a perfectly realistic *S* who succeeds on a constant proportion of trials will, in the long run, expect (predict) success on that proportion of trials. He will show *probability-matching*, in the terminology of the work on probability learning and guessing behavior. An attempt is made to support this proposal experimentally by showing that *S*s do match probabilities in predicting the successes of another, unknown person.

However, one of the consistent findings in studies of levels of aspiration has been that people usually set their level of aspiration somewhat higher than their previous level of achievement; they show positive goal discrepancies. By analogy, it is hypothesized that *S*s will in some degree overpredict their personal successes, i.e., will predict successes more often than they achieve them, rather than match probabilities.

METHOD

The tasks given *S*s in this experiment are best described by the instructions given them.

Predictions of another's successes.—The instructions given *S*s in this control condition follow:

"This is a study of how well people can judge other people's successes and failures. You are to judge the results of a word association experiment that was recently taken by another student.

"One by one, he was given a series of words. He tried to give a correct association for each of these words according to rules which he was not told. He was told, after he gave each association, whether it was correct or incorrect. In other words, by noting whether his response was correct or not, he tried to learn the rules of association as he worked through the series of words.

"I will read to you the same series of words, in the same order, together with his associations. After each pair, I want you to tell me whether you think his association was correct or incorrect according to the rules (which I will not tell you, either). The rules remain the same throughout the experiment, and you may be able to discover some of them as we work through the series.

"Here are the steps we will go through:

1. I will read the word and his association, pair by pair.
2. After each pair, you are to say whether you think his association was correct or incorrect.
3. After your judgment, I will tell you whether his association was correct or incorrect."

The task was performed just as described in the instructions, *E* reading an actual transcript of the associations given by one of the experimental *S*s. There were no actual rules for associating words; *S* was told that the other *S* had succeeded or failed according to a programmed sequence which included 10 successes in each of five 30-trial blocks but was otherwise random.

Predictions of personal successes.—The instructions given *S*s in the experimental conditions follow:

"This is a word association experiment. On each of the cards in front of you is written one word. Certain other words can be correctly associated with each of these words, according to rules which I will not tell you. The rules will remain the same throughout the experiment, and you may be able to discover some of them as you work.

"You are to give one word in response to the word on each card. Your response should be a word you think may be correctly associated with the word on the card according to the rules. Before you give each response I want you to say whether you *expect* (not *hope*) it to be correct or incorrect. After you give your response I will tell you whether it actually was a correct or an incorrect one according to the rules.

"You should work for two goals: (*a*) to give correct responses whenever possible; (*b*) to judge as accurately as possible in each case whether your response actually will be correct or incorrect.

"These are the steps to go through for each card:

1. Pick up the card, turn it over, and read the word on the card *out loud*.
2. Decide on a word (*one* word) to give in response to the word on the card.
3. Say, 'I expect this to be correct,' or 'I expect this to be incorrect,' according to your best judgment.
4. Say your response word. (I will then say, 'That's correct,' if your response is a correct association according to the rules, or 'That's incorrect,' if it is not.)
5. Lay the card face down in a new pile. Pick up the next card and repeat the above steps."

The stimulus material consisted of 151 3 × 5-in. white cards on each of which was typed a single English word. The words used may be described by the following random sample: *because, comfort, gone, ink, nine, police, smooth, surface, true, why.*

The task was performed just as described in the instructions, except that there were no actual rules for associating words. The *E* told *S* he had succeeded or failed (had given a correct or incorrect response) according to a programmed sequence.

Each *S* controlled the pacing of his trials. The mean time taken to complete 151 trials was 32.5 min.; the range was 15 to 60 min. Time taken to complete the experiment was not related to the experimental conditions which will be discussed or to individual differences on other scores.

Five sequences of successes and failures were programmed. The essential difference between them was the proportion of successes, which was $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, or $\frac{5}{6}$ in the various cases. These proportions held true exactly within 30-trial blocks of each sequence; that is, in successive 30-trial blocks *S* had 5, 10, 15, 20, or 25 successes, the same number in each block according to the sequence *E* was giving him. Trial 151 was a success in all sequences. Subject to these restrictions, the sequences of successes and failures were random.

The experimental conditions varied only in which of the five sequences of successes and failures was administered. In the rest of this article, they will be referred to as the $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, and $\frac{5}{6}$ conditions.

Subjects.—The *S*s were 96 male undergraduates from an introductory psychology course at Johns Hopkins, 16 in each condition.

RESULTS

Predictions of another's success.—Table 1 shows by 30-trial blocks the mean number of trials on which *S*s in the control condition predicted successes, together with the standard deviation of their scores for each block. These *S*s were making predictions for another, unknown person who actually succeeded on 10 of the 30 trials in each block. For a while, *S*s predicted somewhat more suc-

TABLE 1

Predictions of Another's Successes

Trials	Mean	*SD*
1–30	13.6	3.5
31–60	13.9	3.9
61–90	11.9	4.8
91–120	10.7	3.8
121–150	9.6	3.1

cesses for him than he attained, but in the latter blocks they predicted for him about the number he actually attained, showing probability-matching.

Predictions of personal successes.—Table 2 shows corresponding data for each of the experimental conditions. These *S*s were making predictions of their own successes while actually succeeding on 5, 10, 15, 20, or 25 trials in each 30-trial block.

The *S*s did not optimistically predict success all the time at any stage in any condition, even in the condition in which they were succeeding on $\frac{5}{6}$ of the trials. They predicted success on a proportion of trials which was positively related to the proportion of trials on which they actually succeeded—the more successes, the more predictions of successes. This result is analogous to the usual finding that level of aspiration is an increasing function of level of previous achievement. And it is like the results of probability-learning experiments in general in that frequency of prediction of an event is always an increasing function of the frequency of its occurrence.

In most probability-learning experiments, however, *S*s approximate probability-matching over many trials, just as they did in predicting another's successes in the present experiment. The *S*s did not show probability-matching in predicting their own successes. If probability-matching is taken as a baseline, *S*s persistently overpredicted success in most of these conditions; they predicted it far more often than they obtained it. This overprediction was greatest for the least successful conditions, lessening as the proportion of actual successes increased, until in the $\frac{5}{6}$ condition *S*s actually slightly underpredicted success.

Predictions of success in the $\frac{1}{6}$ condition show a decided drop over trials, although even in the last blocks *S*s on the average predict almost twice as many successes as they obtained.

Individual differences as reflected in the standard deviations in Table 2 are somewhat larger than in typical probability-learning experiments (Estes & Straughan, 1954), but it was rare for *S*s to approach the extremes of predicting all successes or all failures. Only one *S* in the entire experiment gave the same prediction on all 151 trials, an *S* in the $\frac{1}{6}$ condi-

TABLE 2

Predictions of Personal Successes

Trials	Cond. $\frac{1}{6}$		Cond. $\frac{1}{3}$		Cond. $\frac{1}{2}$		Cond. $\frac{2}{3}$		Cond. $\frac{5}{6}$	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1–30	14.9	7.0	17.6	4.2	17.6	4.7	23.1	5.4	23.8	4.4
31–60	13.4	6.7	17.8	3.2	19.2	5.8	21.7	6.5	23.9	4.7
61–90	10.4	7.5	16.6	5.7	20.4	5.1	22.6	5.7	23.4	6.0
91–120	9.8	7.2	16.2	5.3	17.9	7.1	22.1	6.1	23.9	5.1
121–150	9.1	6.9	19.4	4.1	19.2	5.7	22.3	6.4	25.5	4.9

tion who predicted failures throughout. During the last 30-trial block, four *S*s in the $\frac{5}{6}$ condition and one *S* in the $\frac{1}{2}$ condition predicted success on every trial.

Mean transition probabilities.—Some empirical transition probabilities were obtained for the $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ conditions. For each of the 48 *S*s in these conditions, the following proportions were calculated.

$P(C_S|C_S,C_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be correct and then been told by *E* that he was indeed correct.

$P(C_S|C_S,I_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be correct but then been told by *E* that he was incorrect.

$P(C_S|I_S,C_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be incorrect but then been told by *E* that he was correct.

$P(C_S|I_S,I_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be incorrect and then been told by *E* that he was indeed incorrect.

These proportions were averaged over the 48 *S*s to yield four mean transition probabilities, which are as follows:

$$P(C_S|C_S,C_E) = .69.$$
$$P(C_S|C_S,I_E) = .62.$$
$$P(C_S|I_S,C_E) = .63.$$
$$P(C_S|I_S,I_E) = .65.$$

A Friedman analysis of variance by ranks was performed on the four sets of proportions, yielding a χ^2_r of 12.73, significant at the .01 level with 3 *df*. This result justifies a conclusion that at least the larger differences among the figures above are reliable.

A useful aid to interpreting these figures is to compare the averages of pairs of them. There are three ways this can be done. One is to compare the first two with the last two. The average of $P(C_S|C_S,C_E)$ and $P(C_S|C_S,I_E)$ is .655. The average of $P(C_S|I_S,C_E)$ and $P(C_S|I_S,I_E)$ is .64. This comparison indicates *S*s, on the average, are slightly perseverative in their predictions of correct and incorrect. They tend to predict success on a given trial if they predicted success on the preceding trial; they tend to predict failure on a given trial if they predicted failure on the preceding trial.

Another comparison is the first and third figures vs. the second and fourth. The average of $P(C_S|C_S,C_E)$ and $P(C_S|I_S,C_E)$ is .66. The average of $P(C_S|C_S,I_E)$ and $P(C_S|I_S,I_E)$ is .64. This comparison indicates *S*s, on the average, tend to predict on a given trial the level of achievement, success or failure, that they attained on the preceding trial.

The final comparison is the first and fourth figures vs. the second and third. The average of $P(C_S|C_S,C_E)$ and $P(C_S|I_S,I_E)$ is .67. The average of $P(C_S|C_S,I_E)$ and $P(C_S|I_S,C_E)$ is .625. This comparison, yielding the strongest contrast of the three, indicates that *S*s tend to predict success on a given trial if their prediction of success or failure on the preceding trial was *accurate*.

Ss' reports on their hypotheses.—After the experiment, *S*s were asked if they had discovered the rules by which words could correctly be associated with the words on the cards. Of the 80 *S*s only one failed to report a hypothesis about the rules. Two *S*s directed their hypotheses to the actual sequence of successes and failures, while all the rest expressed hypotheses about the rules of word associations, hypotheses which were necessarily vague and qualified.

DISCUSSION

The fact that *S*s showed probability-matching in predicting the successes of another, unknown person strongly supports the proposal that a perfectly realistic *S* would show probability-matching in predicting his own successes. Also, the data generally support the hypothesis that *S*s, in actuality, tend to overpredict their personal successes. However, the degree of overprediction is surprisingly great in the least successful conditions and surprisingly slight in the most successful conditions.

It seems worthwhile to see how well the lessening of overprediction with increasing success can be accounted for by the traditional explanatory concepts of level-of-aspiration research. It is generally accepted that what makes people depart from the best realism of which they are capable in setting levels of aspiration is the varying attractiveness (or valence or utility) of varying levels of achievement. The greater attractiveness of higher, more difficult levels makes a person aspire to a level somewhat above that which past experience and other information objectively would dictate, the size of the so-called positive goal discrepancy depending on situational requirements for realism (Festinger, 1942). In the case of a two-level achievement scale, it would seem that the perceived difficulty, hence attractiveness, of achieving successes must depend primarily on the relative frequency with which *S*s achieve them. The fewer the successes *S* achieves, the more worthwhile and attractive they are, and the stronger is his tendency to depart from realistic probability matching to overpredict them. This formulation ignores the possibility that as the task becomes more difficult and success becomes more attractive, failure may also become more acceptable, less unattractive, but it is probable that the attractiveness of failure is less labile than that of success and is therefore a minor consideration.

The mean transition probabilities reported indicate that average recency effects were rather slight, which was to be expected from other probability-learning studies in which *S*s' bases for predictions were directed away from the sequence of events itself to other factors determining the events (Goodnow, 1955). The evidence for at least three recency effects, however slight, does indicate that detailed specification of *S*s' predictions would require consideration of many determinants. It is likely too that there are individual differences in such recency effects. For example, there were indications in the data that degree of overprediction of successes was associated with degree of response perseveration.

It should be emphasized that the observed large departures from the supposed realism of probability matching were obtained despite admonitions to *S*s in the instructions to be realistic. It is known that *S*s given such admonitions when setting levels of aspiration on a many-valued achievement scale show only slight goal discrepancies (Festinger, 1942). Apparently such realism is not so easily obtained on a two-category achievement scale. Now the work on probability-learning has established that people can discriminate probabilities in the sense that they predict events with about the same frequency as they occur. They have the discriminative capacity to be realistic in predicting successes on a two-level scale, just as on a many-level scale, but are not so easily compelled to be realistic, perhaps because their lack of realism is less conspicuous to themselves and to others. It appears that in promoting realism in people's aspirations, endeavors, and self-evaluations, it is very important to provide them with feedback information more detailed than mere word that they have succeeded or failed.

SUMMARY

This experiment was concerned with *S*s' predictions of successes and failures on a word-association task where the actual sequence of successes and failures had been programmed by *E*.

In predicting the successes of another, unknown, person who succeeded on $\frac{1}{3}$ of the trials, *S*s eventually showed probability matching, predicting successes at about the frequency with which they occurred.

In predicting their personal successes on the task, the proportion of trials on which *S*s predicted success was an increasing function of the proportion of trials on which they actually succeeded, but they persistently predicted more successes than they actually obtained. This overprediction of successes was considered to be analogous to positive goal discrepancies in the setting of levels of aspiration on a many-valued achievement scale. Since *S*s were admonished to be realistic in their predictions, it was concluded that realism in predictions of personal successes and failures (identified as probability matching) is not as easily obtained as realism in the setting of levels of aspiration on a more finely differentiated achievement scale. The higher the proportion of trials on which *S*s were successful, however, the less they overpredicted, and *S*s who succeeded on $\frac{5}{6}$ of the trials did not overpredict successes.

Examination of first-order transition probabilities in their predictions showed that *S*s were more likely to predict success if they had succeeded on the previous trial than if they had not, they were more likely to predict success if they had predicted success on the previous trial than if they had not, and they were more likely to predict success if they had predicted accurately on the preceding trial than if they had not.

REFERENCES

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

FESTINGER, L. Wish, expectation, and group standards as factors influencing level of aspiration. *J. abnorm. soc. Psychol.*, 1942, **37**, 184–200.

GOODNOW, J. J. Response-sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 624–630.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. I. New York: Ronald, 1944. Pp. 333–378.

SIMON, J. R., SHAW, M. E., & GILCHRIST, J. C. Some effects of prearranged performance scores upon the level of aspiration. *J. exp. Psychol.*, 1954, **47**, 10–12.

(Received February 12, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

GRADIENTS OF GENERALIZATION IN SECONDARY REINFORCEMENT

BRUCE O. BERGUM[1]

Northwestern University

Studies concerned with the generalization of secondary reinforcement (Denny, 1948; Ehrenfreund: 1949, 1954; Saltzman, 1950) typically involve comparison of two groups, one with similar, the other with less similar, secondary reinforcing cues. These studies have established the fact that stimulus generalization is a variable in secondary reinforcement, but they do not provide evidence as to whether there is a gradient in such generalization. The purpose of the present study is to determine whether such a gradient exists and, if one does, to determine its general form.

The present study is modeled somewhat after those of Saltzman (1949; 1950). Basically, it involves training in a runway to establish the secondary reinforcing value of a given stimulus, followed by testing for generalization in a new learning situation (**U** maze) with five physically defined levels of similarity to the original training stimulus.

METHOD

Subjects.—The *S*s were 100 male albino rats of the Sprague-Dawley strain, ranging in age from 60 to 75 days.

Apparatus.—The apparatus consisted of a runway, a single-choice **U** maze, an interchangeable starting box, one black and one white goal box for use with the runway, and six goal boxes for the maze. All of the apparatus (except the starting box which was 9-in. square) was 6 in. wide and 8 in. deep in interior dimensions.

The starting box, runway, and **U** maze were painted with alternating ½-in. wide black and white vertical stripes to make them distinctive from the goal boxes, and to minimize generalization from the goal boxes to the various alleys.

The six maze goal boxes varied in shade from white through black in approximately subjectively equal intensity intervals as determined by three human judges. Illuminometer readings in foot-candles for the six shades *in situ* were: 29.2 for the white, 14.5 for the light gray, 6.8 for the medium light gray, 3.6 for the medium dark gray, 1.9 for the dark gray, and 1.0 for the black.

Procedure.—For six days prior to training all *S*s were placed on a feeding schedule consisting of 10 gm. of dry pellet food in the home cage for 1 hr. in every 24. During this preliminary period *S*s were tamed and familiarized with both the dry pellets and the various parts of the apparatus. The reward pellets used in the runway were standard .05 gm. laboratory pellets.

On Days 1 and 2 of the experiment proper, each *S* received 40 trials in the runway on each day. On the first trial each day *S* ran to a black-positive (white for half the animals) goal box where he received a single pellet of food. Following this trial, *S* received each day an additional 24 reinforced trials running to the positive goal box, randomly alternated with seven nonreinforced trials to the positive goal box and eight nonreinforced trials to the negative goal box. At the completion of a run *S* was allowed to remain in the goal box for 10 sec., after which he was placed in the home cage for 15 sec. preceding the next trial.

On Day 3 of training *S*s were randomly divided into five groups for training in the **U** maze. For the 20 *S*s in Group I the "correct" (i.e., secondarily rewarded) maze goal box was the same color as the box in which *S* was reinforced during runway training. For Group II the correct maze goal box was the one most similar in color to the reinforced runway box. For Group III the correct maze box was that second most similar, for Group IV the correct box was that third most similar, for Group V the correct box was that least similar to the positive runway goal box. Therefore, for Group V the correct maze goal box was most similar, but not identical, to the negative runway box. In addition, the various counterbalancing procedures were such as to permit the testing of gradients both from white to black and from black to white.

---

[1] Based upon a dissertation submitted to the Graduate School, Northwestern University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to express gratitude to his adviser, Carl P. Duncan, and to Donald J. Lewis and John W. Cotton for their assistance in carrying out the study.

For half of the *S*s the correct goal box appeared on the left side of the maze, and for the remaining half, on the right. For all *S*s the "incorrect" maze goal box was the never reinforced (black or white) goal box used in the runway on the previous two days.

All *S*s were given one forced trial to the correct goal box just preceding the first test trial in the **U** maze, and all *S*s then received 15 trials in the **U** maze with a modified correction technique; i.e., *S*s were allowed to correct the turn they made at the choice point as long as they did not enter a goal box. However, once they entered either goal box the door was dropped behind them and they were kept in the box for 10 sec. The *S*s received *no* primary reward in either goal box at any time during the maze training. Following each trial *S* was placed in the home cage for 20 sec.

The response measures during maze training were (*a*) reciprocal running times measured from the time the door was raised on the starting box until *S* made the *first* turn at the choice point on each trial, (*b*) frequency with which the first turn was correct, and (*c*) frequency with which the goal box entry was correct.

RESULTS

Analyses of the maze data will be presented for each of the three response measures separately.

Number of correct turns.—The score for each *S* for this measure was the total number of correct turns made during all 15 trials in the maze. The mean number of correct turns for each of the 10 groups ($N = 10$ per group) is shown in Table 1. In Table 1 and in subsequent similar tables "reinforced color" indicates the color of the positive goal box during runway training, "similarity level" indicates degree of similarity of the correct maze goal box to the reinforced color.

TABLE 1

MEAN NUMBER OF CORRECT TURNS OVER ALL 15 TRIALS IN THE **U** MAZE

Reinforced Color	Similarity Level					Mean
	I	II	III	IV	V	
White	9.90	9.40	7.90	8.10	7.90	8.64
Black	10.50	8.80	9.10	7.80	6.60	8.56
Mean	10.20	9.10	8.50	7.95	7.25	

The grand means in the bottom row of Table 1 show a systematic trend without reversal in the effects of the similarity variable with the most similar **U**-maze goal box yielding the greatest number of correct turns and the least similar goal box yielding the smallest number. The color of the originally reinforced runway goal box does not appear to have differentially affected performance.

The data were analyzed by a 2 × 5 factorial analysis of variance in which the two dimensions were (*a*) reinforced color, and (*b*) similarity level. The results of the analysis, presented in Table 2, show that the similarity variable was highly significant. Bartlett's test for homogeneity of variance yielded a χ^2 of 12.99 for 9 *df*, which is not significant.

The systematic effect of the similarity variable is shown by the lower curve in Fig. 1, in which the mean

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT TURNS

Source	*df*	*MS*	*F*
Reinforced color	1	0.160	
Similarity	4	25.325	7.43*
C × S	4	4.885	1.43
Within	90	3.411	

* $P < .001$.

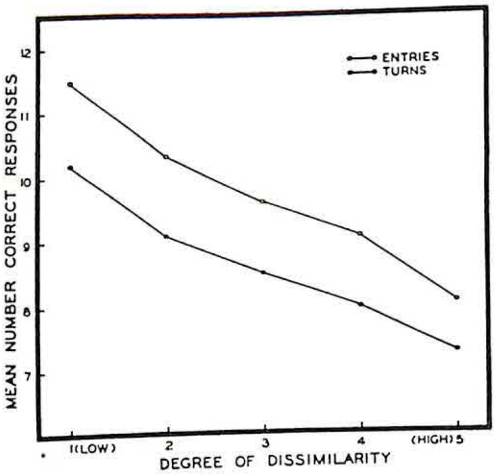


FIG. 1. Mean correct turns and mean correct goal box entries in the maze as a function of degree of dissimilarity in brightness of the reinforced runway goal box to the correct maze goal box.

number of correct turns is plotted against similarity level. With the exception of a slight depression at Level V below the chance level of performance (7.5 mean correct turns), the empirical curve is very regular. In Fig. 1 the similarity levels are equally spaced on the baseline for human judges. If the baseline were plotted in log foot-candles of goal box luminance, the curves would look much the same.

To examine the possible improvement in performance through the course of experience in the maze, the total numbers of correct turns during the first five, second five, and last five trials were determined for each S at each similarity level. All similarity groups appear to show improvement in performance with practice most of the change occurring in the first 10 trials. The differential effects of the similarity variable appear to establish and maintain themselves almost from the beginning of testing.

TABLE 3

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT TURNS AT VARIOUS STAGES OF PRACTICE

Source	*df*	*MS*	*F*
Similarity	4	0.835	41.75*
Stages	2	1.190	59.50*
Sim. × Stage	8	0.020	

* $P < .01$.

In order to determine whether the apparent change in performance shown in the center curve of Fig. 2 was significant, an analysis of variance for similarity level and stage of practice, summarized in Table 3, was performed on the 15 means. Table 3 shows that, as before, the similarity effects were significant beyond the 1% level. Of more immediate interest, the stages-of-practice variable was also significant beyond the 1% level indicating a significant improvement in performance in the course of U-maze training.

Number of correct goal box entries.—On this measure each S's score consisted of the total number of correct goal box entries made during all 15 trials. The mean number of entries is presented in Table 4. These means, also shown in the upper curve of Fig. 1, parallel almost exactly those for the turns data except that they tend to average slightly more than one response higher.

The entries data were analyzed by the same type of analysis of variance

TABLE 4

MEAN NUMBER OF CORRECT ENTRIES OVER ALL 15 TRIALS IN THE U MAZE

Reinforced Color	Similarity Level					Mean
	I	II	III	IV	V	
White	11.60	10.30	9.70	9.20	8.60	9.88
Black	11.40	10.40	9.50	8.90	7.40	9.52
Mean	11.50	10.35	9.60	9.05	8.00	

used on the turns data (Table 2). No table is presented for this analysis since the results were almost identical to those for turns. Neither original reinforced color nor the interaction was significant. Similarity was significant beyond the .001 level ($F = 8.16$ for 4 and 90 *df*). The Bartlett test for homogeneity of variance was not significant ($\chi^2 = 4.49$ for 9 *df*).

The top curve of Fig. 2 shows the means for the entries data broken down into thirds of practice. The results are much the same as for the turns data except that in this case the means tend to fall at or above chance, probably because an alternation tendency should have less effect on entries than on turns. With the correction procedure used, *S* could still enter the correct goal box even after making an incorrect turn.

The same type of analysis of variance used for the turns data was done on the means for entries. The results were much the same as those shown in summary Table 3; similarity was significant at the 1% level ($F = 17.06$ for 4 and 8 *df*), stages of practice was significant at the 1% level ($F = 10.15$). Thus, a significant improvement in performance also took place in terms of the entries measure.

Except in reflecting a slightly higher level of performance, the results for the number of correct entries parallel exactly those for the turns data. The tendency to yield a higher level of performance may indicate that the entries measure is somewhat less affected by irrelevant turn biases than is the turning response, and probably is a more sensitive measure of secondary reinforcing effects. Both measures, however, apparently demonstrate equally well the general form of the generalization gradient.

Running times.—This measure was scored as the reciprocal of the time required by *S* to leave the starting box and traverse the stem of the maze to the choice point. The raw score for each *S* consisted of the sum of the reciprocal running times over all 15 trials.

The mean reciprocal time for each group is shown in Table 5. These data reflect the same over-all gradient found for the preceding two measures, but the relationship is not quite as smooth in this case.

The same type of analysis of variance (Table 2) used on the other two measures was applied to the reciprocal running times. The results were essentially the same as those for turns and entries. Neither reinforced goal box color ($F < 1.00$) nor the interaction ($F = 1.51$) was significant. Similarity, however, was significant beyond the .001 level ($F = 7.20$). Bartlett's homogeneity of variance test yielded a nonsignificant χ^2 of 15.18 for 9 *df*.

As in the case of turns and entries, the running time data were broken down into three stages of training, as shown again in Fig. 2. Of particular interest is the fact that, in contrast to the progressive improvement shown by the turns and entries data, running times display a typical extinction curve form.

TABLE 5

MEAN RECIPROCAL RUNNING TIMES OVER ALL 15 TRIALS IN THE U MAZE

Reinforced Color	Similarity Level					Mean
	I	II	III	IV	V	
White	10.35	11.80	9.28	8.35	7.28	9.41
Black	13.17	12.37	5.99	8.39	6.19	9.22
Mean	11.76	12.08	7.63	8.37	6.73	

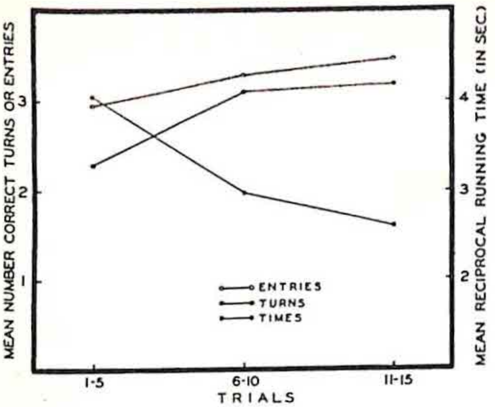


FIG. 2. Mean correct turns, mean correct goal box entries, and mean reciprocal running times in terms of three stages of practice in the maze.

The same type of analysis of variance used to analyze stages of practice in turns and entries data was performed on the 15 means for similarity and stage of practice. As expected, the similarity effects were significant beyond the 1% level ($F = 7.16$ for 4 and 8 *df*). The F for stages of training was 11.06, which for 2 and 8 *df* is significant beyond the 1% level. Thus, it appears that significant extinction of the running response occurred in the course of U-maze training. This presents the interesting paradox of *S*s running slower and slower while at the same time showing significant improvement in the performance of correct turns and entries. This relationship is shown graphically in Fig. 2.

Except for showing a decrease in performance level, the evidence from the running times data follows the same general trend as that found with the turns and entries data. The conclusion that secondary reinforcing stimuli generalize systematically as a function of their objective similarity to the originally reinforced stimulus is strongly supported.

DISCUSSION

The purpose of this experiment was to study the generalization of secondary reinforcement as a function of stimulus similarity along an intensity continuum, and to determine the general form of the gradients associated with three response measures.

The results of the study were unambiguous. Response strength was directly related to the degree of similarity of the secondary reinforcing stimulus to the originally reinforced stimulus. These results are in conformity with what would be expected on the basis of a large body of evidence relating to stimulus generalization. In terms of gradient forms, both turns and entries yielded approximately straight-line gradients, but while end-point differences for the running time measures were significant, the form of the gradient was not as clear.

The significant improvement in performance for the turns and entries data is open to alternative interpretations as a result of the specific U-maze procedure employed in this study. Since the procedure involved a preliminary forced trial to the correct side, it is possible that this may have led to an increased probability of an incorrect turn on the first test trial. The slight depression below the chance level of 2.5 correct turns in the first five trials, apparent in Fig. 2, lends some support to this notion. The improved performance over trials could thus simply represent recovery from this alternation tendency. Dember and Fowler (1958) present some of the arguments for the alternation interpretation of performance increments and a summary of the relevant literature.

The alternative interpretation would be that the improved performance over trials found in the present study represents learning. If the improvement is so interpreted, the results for all three response measures might be explained in terms of the more recent theoretical writings of Spence (1956). In his formulation, N, number of S-R occurrences, becomes the single effective variable determining associative strength, while an indeterminate number of other variables, including N, affect the intervening variable K, a motivational factor. Secdary reinforcement is the association of r_g's with stimuli proximate to the primary reinforcer by a process of classical conditioning. Elicitation of these fractional responses results in heightened tension which contributes to the general drive level. Thus, secondary reinforcement is a motivational variable consisting of classically conditioned r_g's which follow the traditional laws of conditioning.

This formulation would lead to certain predictions in the context of the present study. Thus, if r_g is classically conditioned, we would expect that in the course of U-maze training, where primary reinforcement never occurred, r_g would tend to extinguish, motivation would decrease, and S would tend to run more and more slowly in the course of training. As indicated in Fig. 2, this prediction was borne out.

Presumably, r_g is elicited to a greater or lesser extent in the correct goal box of the U maze, and as practice continues, stimuli at progressively earlier points come to elicit r_g. Elicitation of these r_g's would increase K and thus increase the motivational level of S. Given the secondary reinforcing effects of having entered the "correct" goal box at least once, the difference in H between "correct" and "incorrect" habits would be multiplied by a factor greater than unity, increasing the probability of correct turns. Differences in the probability of eliciting r_g in the goal box as a function of similarity to the originally reinforced runway goal box would account both for the over-all differences in running times among groups and the differences in turns. With lower probabilities of eliciting r_g, less K would be generated and would result in a decrease in the probability of a correct turn. The effect of this mechanism would, of course, be greatest early in learning, before it had extinguished, with the effect leveling off as r_g became progressively less functional.

Finally, given a correct turn, and once in sight of the correct goal box, S's drive level would be increased. This would probably lead to an increase in the speed of locomotion toward the box, but in any case should lead to its eventual entry. In the case of the negative goal box, the drive would fail to be increased, thus resulting in a relative decrease in the probability of S's entering the correct goal box. Again, the evidence in Fig. 2 supports this prediction.

SUMMARY

The experiment was designed to study the generalization of secondary reinforcement as a function of stimulus similarity along an intensity continuum, and to determine the general form of the gradients associated with three response measures.

Five groups of white rats (20 per group) received 80 trials in a runway on two consecutive days, running to a partially reinforced goal box on 64 trials and to a nonreinforced goal box on 16 interspersed trials. The goal boxes were black or white. On the third day of training all *Ss* were given 15 trials in a U maze in which the nonreinforced goal box appeared at the end of one arm and a goal box bearing one of five degrees of similarity in brightness to the originally reinforced goal box, appeared at the end of the other arm. The points along the similarity continuum differed by equal appearing intervals according to human judges. The *Ss* were never given food reinforcement in either goal box in the U maze.

Approximately straight-line generalization gradients were shown by two of the three response measures, correct turns, and correct goal box entries, and these gradients yielded statistically significant end-point differences in all cases. In the case of reciprocal running times the end-point differences were significant, but the form of gradient was not as clear. There was no difference between generalizing from white to black and black to white. The data were further analyzed in terms of stage-of-practice in the U maze and it was found that turns and entries showed significant improvements in performance, while simultaneously the running time showed a significant decrease in performance level.

REFERENCES

DEMBER, W. N., & FOWLER, H. Spontaneous alternation behavior. *Psychol. Bull.*, 1958, **55**, 412–428.

DENNY, M. R. The effect of using differential end boxes in a simple T-maze learning situation. *J. exp. Psychol.*, 1948, **38**, 245–249.

EHRENFREUND, D. Effect of a secondary reinforcing agent in black-white discrimination. *J. comp. physiol. Psychol.*, 1949, **42**, 1–5.

EHRENFREUND, D. Generalization of secondary reinforcement in discrimination learning. *J. comp. physiol. Psychol.*, 1954, **47**, 311–314.

SALTZMAN, I. J. Maze learning in the absence of primary reinforcement: A study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, **42**, 161–173.

SALTZMAN, I. J. Generalization of secondary reinforcement. *J. exp. Psychol.*, 1950, **40**, 189–193.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received February 15, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

RETENTION OF RESPONSES TO STIMULUS CLASSES AND TO SPECIFIC STIMULI[1]

KENNETH E. LLOYD[2]

Ohio State University

It is frequently assumed that retention of meaningful material is superior to that of nonmeaningful material. While precise definitions of meaningfulness are elusive, it is often stated that prose, concepts, or substance material meet the criteria of meaningfulness. Comparisons of retention of such materials with that of nonsense syllables have been made, but rather than specifying the measurable variables that contribute to the observed retention differences, explanations have centered around the qualitative aspects of the learning materials. One variable that could influence such results is the number of response units that must be made by *S*s who learn different kinds of material.

For example, Reed (1946) concluded that retention of concepts was far superior to retention of nonsense syllables. The *S*s in his concept group associated six nonsense syllable responses to 42 verbal stimuli which could be grouped into six concepts with seven instances of each. The retention of these *S*s was compared with that of Ebbinghaus (1913) who usually learned eight lists of 13 syllables in one session; i.e., the retention of six responses was compared with that of 104 responses. Katona (1940) requested *S*s to learn a series of 24 numbers, but varied his instructions to different groups in such a way that discovery of a rule, by some of the groups, reduced the number of separate items that had to be remembered from 24 to 4. The superior retention of these concept groups may have been a function of the relatively fewer number of responses learned by them, rather than a function of the meaningfulness or organization of the material. The present experiment controlled this possibility by equating the number of response units which *S*s learned.

The hypothesis tested was that there would be no difference in retention between *S*s who acquired responses to specific stimuli and *S*s who acquired responses to classes of stimuli if the number of response units learned was the same in both situations. Responding to specific stimuli means that a given response is emitted only in the presence of a given stimulus; responding to classes of stimuli means that a given response is emitted in the presence of any instance of a class of stimuli.

No presumption is made that the latter condition is necessarily more meaningful than the former. Nevertheless, responding to classes of stimuli should involve a mediating response while responding to specific stimuli need not. It is possible to determine empirically whether or not *S*s in the class groups are employing a mediating word. If *S*s who learn arbitrary responses to instances of a stimulus class are presented with different instances of this class, their performance should exceed that of *S*s who have learned responses only to specific words. When the new instance is presented, *S* makes the mediating response which then sets the occasion for the arbitrary response. If the response has been made only to a specific stimulus, then no mediating response need be employed, and *S* is not set for a new instance.

---

[1] This article is based on a dissertation submitted to the Graduate School of the Ohio State University in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his appreciation to Paul M. Fitts whose stimulation greatly facilitated this work.

[2] The author is now at Washington State University.

For example, one group could learn R_1 to the words *priest*, *monk*, and *deacon*, while another group learned the same response to the one word *priest*. If both groups are later presented with the word *bishop*, then the former group should emit R_1 more quickly than the latter. Positive transfer effects would undoubtedly operate in both conditions, but the class groups should benefit more if they have employed a mediating word such as *churchmen*.

A transfer task was used in this study as a means of testing the proposition that one group responded to classes of stimuli while the other group responded to specific stimuli. No retention differences were expected, however, since the number of responses in the two conditions was identical.

METHOD

Design.—The design consisted of three steps, an original learning (L-I), a recall trial, and a transfer task (L-II). During L-I two specific groups (experimental and control), and two class groups (experimental and control) learned eight number responses either to eight specific words, or to 24 words which could be categorized into eight classes, i.e., three instances of each class. After the recall trial, the experimental specific group (Group ES) learned a second list containing eight different specific words from the same classes used in L-I. The control specific group (Group CS) transferred to eight specific words from entirely different classes. The experimental class group (Group EC) and its control (Group CC) transferred to lists containing three new instances of the old classes, and of new classes, respectively. The same eight responses were used in both L-I and L-II. Table 1 illustrates the design with words taken from two classes of stimuli.

TABLE 1

STIMULUS WORDS AND THEIR RESPONSE NUMBERS FOR THE FOUR GROUPS DURING THE ORIGINAL LEARNING AND THE TRANSFER TASKS (THE WORDS ARE INSTANCES OF TWO CLASSES)

Group	Learn I		Learn II	
	Stim.	Resp.	Stim.	Resp.
ES	Priest[a]	1	Bishop	1
CS	Bishop[a]	1	Cottage	1
EC	Priest Monk Deacon	1 1 1	Bishop Missionary Chaplain	1 1 1
CC	Bishop Missionary Chaplain	1 1 1	Cottage Garage Skyscraper	1 1 1

[a] *S*s in Groups ES and CS were divided into six subgroups so that all the instances of a class could serve as specific words; e.g., during L-I seven *S*s in Group ES saw only the word *priest*, seven other *S*s saw only the word *monk*, etc. This counterbalancing produced no performance differences.

Procedure.—The 72 stimulus words were selected on the basis of their class membership and their frequency of use (Thorndike & Lorge, 1952). Within each class there were both relatively frequent and relatively infrequent words. Eight preliminary *S*s were asked to sort these words plus 72 control words into categories which they deemed appropriate. Only 12 of the 576 experimental words sorted were misplaced, indicating high agreement as to the appropriate class for each word.

A trial was defined as the presentation of 24 S-R pairs. The same eight S-R pairs were repeated three times within a trial for the specific groups. For the class groups the same eight responses were repeated three times, but were paired with a different stimulus word each time they occurred within a trial.

The word and number pairs were presented to *S*s in a typical paired-associates manner in a Patterson-type memory drum at a 2:2-sec. rate with a 24-sec. intertrial interval. Two different presentation orders were used. The *S*s responded into a microphone which was part of a voice key used to obtain response latencies in addition to the usual correct response measure. Before beginning L-I all *S*s were informed that the responses would be the numbers one through eight, and were shown a list of the 72 words grouped into the 16 classes, but *S* did not know which number would be paired with which word, or class of words. One week later *S*s returned for the recall trial, and L-II.

Subjects.—The 21 *S*s in each group were elementary psychology students of both sexes fulfilling their subject-hour requirement at the Ohio State University.

Since responding to classes of stimuli was suspected to be more difficult than responding to specific stimuli, two experiments were conducted. In Exp. I the original learning was to a criterion of one perfect trial. In Exp. II, which employed 84 additional *S*s from the same population, L-I was presented for five trials only. The experiments were identical in other respects. Thirty-five additional *S*s completed L-I, but failed to return for the second session.

In summary, *S*s learned the same number of responses to specific words, or to classes of words. At a recall trial no differences between their scores were expected. A transfer task served to determine whether *S*s in the class group were responding to the stimuli as members of classes, and not as specific words.

RESULTS

The mean number of correct responses for all groups is plotted as a function of successive blocks of eight responses for each condition in Fig. 1. Although a trial was defined as 24 responses, there were only eight different responses used in the experiments. Therefore, each trial is a composite of three subtrials; it is these subtrials which are plotted on the abscissa in Fig. 1. Unless noted otherwise, all the statistical analyses were computed for a total trial of 24 responses, however. The *F* ratios computed for all four groups were based on 3 and 80 *df*. The *F* ratios for individual degrees of freedom were based on 1 and 80 *df*.

Table 2 presents the trials to cri-

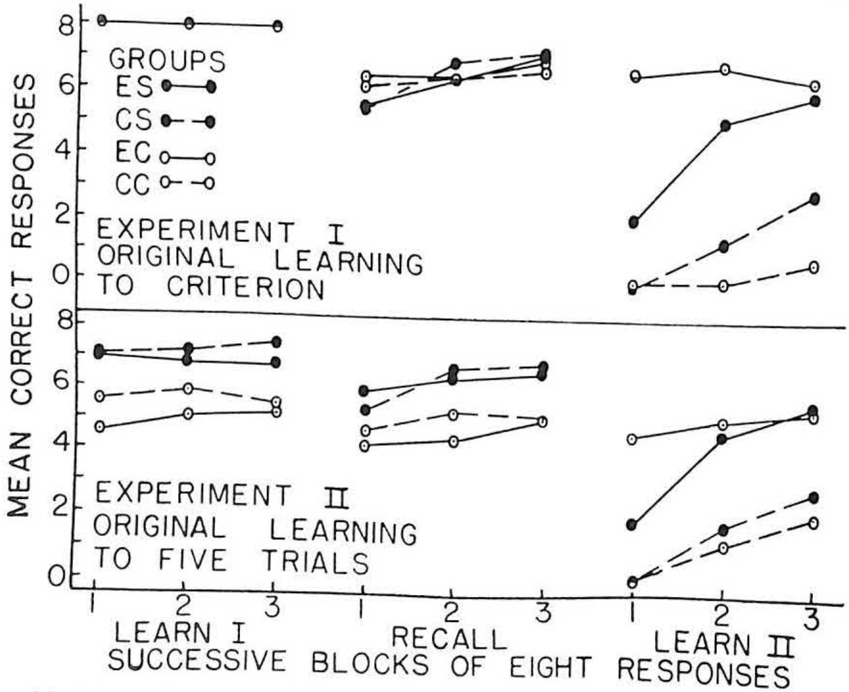


FIG. 1. Mean correct responses for successive blocks of eight responses. Each point on the abscissa represents one subtrial containing all eight responses from a total trial of 24 responses. The last three subtrials of the original learning are plotted at Learn I. All three recall subtrials are plotted at Recall. The first three subtrials of the transfer task are plotted at Learn II.

TABLE 2

Trials to Criterion and Recall in Exp. I

Group	L-I Trials		Recall	
	Mean	*SD*	Mean	*SD*
ES	6.81	2.67	19.19	3.02
CS	7.05	2.05	19.71	4.07
EC	9.14	2.04	20.10	4.70
CC	8.86	2.51	19.43	4.72

TABLE 3

Correct Responses at L-I and Recall in Exp. II

Group	Trial 5, L-I		Recall	
	Mean	*SD*	Mean	*SD*
ES	20.62	3.85	19.27	4.29
CS	21.86	3.13	19.19	4.70
EC	14.86	4.45	13.86	4.03
CC	17.00	4.68	15.48	5.10

terion and recall scores for Exp. I. The F ratio for trials is significant ($F = 5.33$, $P < .01$). Snedecor's (1946) test for individual degrees of freedom indicates that the class groups require more trials to criterion than do the specific groups ($F = 15.76$, $P < .01$). At recall the between-groups variance estimate is less than the within-groups variance, indicating no differences in recall among the groups. A second analysis of variance for the first eight responses yielded another insignificant F ratio.

Table 3 contains the correct responses at L-I (Trial 5) and at recall for Exp. II. Both F ratios are significant ($F = 12.54$, $P < .01$; $F = 7.17$, $P < .01$). By Trial 5 of L-I the specific groups have made more correct responses than the class groups ($F = 33.97$, $P < .01$). At recall the specific groups maintain this superiority ($F = 20.25$, $P < .01$). Neither class nor specific groups differ significantly from each other in any of the above comparisons.

Since the class and specific groups differed in their rates of learning, computation of loss scores (Richardson & Underwood, 1957; Underwood, 1954) provides a more appropriate assessment of retention than do the raw recall scores. Loss scores serve to equate associative strength at the end of learning. They are obtained by subtracting raw recall scores from estimated immediate recall scores which are derived from successive probability curves for each group. In order to obtain an estimated immediate recall score for 24 responses three successive probability curves based on eight responses were computed for each group. The first of these curves indicated the probability of a correct response on the next subtrial following a given number of prior correct responses. The second curve indicated the probability of a correct response on the second subtrial following a given number of prior correct responses, and the third curve provided probabilities for the third subtrial.

Table 4 contains the mean loss scores. Bartlett's test indicated homogeneity of variance in both experiments ($\chi^2 = 4.21$ and 1.46; $df = 3$). Analysis of variance resulted in F ratios of less than one in both experiments. Loss scores were also computed for the first subtrial at recall; again the F ratios in both ex-

TABLE 4

Mean Loss Scores

Group	Exp. I		Exp. II	
	Mean	σ_M	Mean	σ_M
ES	3.95	0.69	2.47	0.71
CS	3.47	0.93	3.36	0.72
EC	2.96	1.06	2.95	0.89
CC	3.76	1.03	3.47	0.75

periments were not significant. The groups did not differ in loss scores in either experiment.

When *S*s do not learn to a criterion, as in Exp. II, some items receive zero reinforcements. The loss scores in Table 4 were obtained by assigning zero probability to such items. Since this method may have underestimated loss scores, a second set of loss scores was obtained by assigning items with zero reinforcements the probability following one reinforcement. This procedure overestimated loss scores. Analysis of variance for the overestimated loss scores was also insignificant ($F = 2.13$, $P > .05$). This F ratio is larger than the F obtained using the first method due to increased loss for the class groups which had more items with zero reinforcements.

The transfer data (L-II, Trial 1) are shown in Table 5. The F ratio in Exp. I was significant ($F = 109.78$, $P < .01$). The test for individual degrees of freedom indicated that the experimental groups differed from the control groups ($F = 285.38$, $P < .01$), that Group EC exceeded ES ($F = 36.41$, $P < .01$), and that Group CS exceeded CC ($F = 7.56$, $P < .01$). In Exp. II the F of 46.31 was again significant ($P < .01$). The rank order of the groups was identical to Exp. I, but the differences were not as great; the experimental groups differed at the 5% level, and the control groups did not differ.

TABLE 5

CORRECT RESPONSE AT L-II, TRIAL 1, IN BOTH EXPERIMENTS

Group	Exp. I		Exp. II	
	Mean	*SD*	Mean	*SD*
ES	13.00	4.38	12.24	5.07
CS	4.24	3.19	4.57	2.56
EC	20.00	4.78	15.29	4.31
CC	1.05	1.29	3.29	2.83

The latency data paralleled the results for correct responses and are not presented. They indicated that responding to classes of stimuli required more time than responding to specific stimuli. However, at L-II in both experiments the latency of Group EC was as short as that of the specific groups. The *S*s in Group EC responded to new instances just as quickly as *S*s in Group ES responded to new specific stimuli.

DISCUSSION

The results of Exp. I indicated no differences in recall when the number of responses was equal for both class and specific groups, and when associative strength at the end of learning was equated. The three successive probability curves for all groups coincided after, at most, eight reinforcements. In the class groups 6% of the items had less than eight reinforcements, compared with 27% in the specific groups. The estimated immediate recall for the class groups was thus largely based on probabilities that did not differ from the probabilities of the specific groups; the same relationship holds for three-fourths of the probabilities contributing to the estimated immediate recall of the specific groups. In addition, the remaining one-fourth of the probabilities for the specific groups, i.e., those based on less than eight reinforcements, were higher than those for the class groups. The over-all effect is to produce equivalent estimates of immediate recall for all groups. When neither immediate recall nor raw recall scores differ, then, of course, the loss scores will not differ.

In Exp. II the class groups learned fewer items than the specific groups in five trials; the former groups also recalled fewer items. The three successive probability curves for all groups again coincided after, at most, eight reinforcements. In contrast to Exp. I, *S*s in the class groups received less than eight reinforcements for 71% of the items compared to 45% for the specific groups. The estimated immediate recall for the class groups was, therefore, largely derived from those parts of the probability curves where the probabilities were lowest. This resulted in lower immediate recall scores relative to the specific groups. The raw recall scores for the class groups were also lower. Again the over-all effect was to produce equivalent loss scores for the class and specific groups.

The transfer data from both experiments strongly suggest that the class groups were responding to the words as if they were members of a class. At the first presentation of a second list containing new instances from the old classes, Group EC was superior to Group ES. It would have been possible for *S*s in Group ES to classify the stimuli during L-I since they had responded to these words many times before coming to the laboratory, but if they had done so, their performance during the first trial of L-II should have equaled that of Group EC. Under such conditions the class groups would ordinarily be predicted to display superior retention; they did not. When the number of responses was held constant and when associative strength at the end of learning was equated, retention did not differ between *S*s who responded to stimuli as members of classes and *S*s who responded to specific stimuli.

SUMMARY

No differences in retention between *S*s responding to classes of stimuli, and *S*s responding to specific stimuli were predicted if the number of responses in both conditions was equal. A transfer task was used to test the expectation that *S*s would respond to the stimuli as instances of classes if they were given an opportunity. When original learning was carried to a criterion specific groups learned at a faster rate than class groups. There were no differences in either raw recall scores or loss scores between the groups. When original learning was for five trials the specific groups made more correct responses, and their raw recall scores were greater than the class groups. Loss scores after incomplete original learning did not differ. The greater transfer effects for the class groups supported the notion that they were responding to the stimulus words as members of classes. Under these conditions retention of responses learned to stimulus classes did not differ from retention of responses to specific stimuli.

REFERENCES

EBBINGHAUS, H. *Memory*. Trans. by H. A. Ruger and C. E. Bussenius. New York: Teachers Coll., Columbia Univer., 1913.

KATONA, G. *Organizing and memorizing*. New York: Columbia Univer. Press, 1940.

REED, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.*, 1946, **36**, 71–87.

RICHARDSON, J., & UNDERWOOD, B. J. Comparing retention of verbal lists after different rates of acquisition. *J. gen. Psychol.*, 1957, **56**, 187–192.

SNEDECOR, C. W. *Statistical methods*. Ames: Iowa State Coll. Press, 1946.

THORNDIKE, E. L., & LORGE, I. *The teachers' word book of 30,000 words*. New York: Teachers Coll., Columbia Univer. Press, 1952.

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.

(Received February 21, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

GROWTH OF SENSATION ON SEVEN CONTINUA AS MEASURED BY FORCE OF HANDGRIP

JOSEPH C. STEVENS, JOEL D. MACK, AND S. S. STEVENS

Harvard University

There is abundant evidence that on prothetic continua the subjective magnitude (ψ) grows approximately as a power function of the stimulus magnitude (ϕ) and that each modality has its characteristic exponent (S. S. Stevens, 1957). The present study is another step in a program designed to show whether functions of this form are able to predict the behavior of observers (*O*s) whose task it is to match the apparent intensity of sensations aroused through different sense modalities. If two continua are governed by power functions,

$$\psi_1 = \phi_1{}^m$$

and

$$\psi_2 = \phi_2{}^n$$

it follows that cross-modality matches, in which ψ_1 is equated to ψ_2, will determine an equal-sensation function of the form,

$$\phi_1 = \phi_2{}^{n/m}.$$

In log-log coordinates this equation determines a straight line with a slope equal to the ratio n/m.

Such equal-sensation functions have been both predicted and obtained by S. S. Stevens (1959a, 1959b) and by J. C. Stevens and Mack (1959). In the present study the forms of five more functions are predicted and to a good approximation verified. The *O*s were asked to produce, on a hand dynamometer, forces that seemed subjectively to match the apparent intensity of various stimulus levels produced by electric shock, white light, white noise, 1000-cycle tone, and vibration on the finger tip. In other words, instead of the usual procedure of asking *O* to estimate numerically the apparent loudness, say, of various stimulus levels, he was asked to express his appreciation of loudness by squeezing. Instead of emitting numbers, as it were, *O* emitted squeezes.

Since the subjective force of handgrip has been shown to grow as the 1.7 power of the physical force exerted (J. C. Stevens & Mack, 1959), the equal-sensation functions ought to have the form,

$$F = I^{n/1.7}$$

where F is force and I is intensity on the continuum of the criterion stimulus to which the apparent force is matched. The exponent n is the exponent governing the growth of subjective magnitude as determined in earlier ratio-scaling experiments, mainly by the method of magnitude estimation. In Table 1 will be found the continua, the approximate value of the exponent (n) for each continuum, the predicted exponent of the equal-sensation function ($n/1.7$), and the exponent actually obtained by matching with handgrip.

APPARATUS AND PROCEDURE

The *O* used his right hand to exert the matching forces on a hand dynamometer. This dynamometer (J. C. Stevens & Mack, 1959), permitted the measurement of force up

---

[1] This research was supported in part by Grant G2668 from the National Science Foundation, and in part by Contract N5onr-1866(15) with the Office of Naval Research (Project Nr142-201, Report PNR-226). Reproduction for any purpose of the United States Government is permitted.

to 100 lb. and was designed to align itself, by means of ball-bearing joints, in the direction of the force applied to its handle. For a full 100-lb. deflection of the dial, it was necessary to displace the handle through a distance of only .039 in.

In each experiment various intensities of a criterion stimulus were presented in a different irregular order to each *O*. The various stimulus levels were then repeated one or more times, again in a different order for each *O*. The *O* was instructed to squeeze the dynamometer with a force that seemed to match the apparent magnitude of the criterion stimulus. He was to do this by "bracketing" and was to signal as precisely as possible the moment at which the matching force was achieved.

Twenty-six different *O*s participated in one or more of the five experiments, 4 in four of the experiments, 3 in three of them, 8 in two of them, and 11 in a single experiment. Three of the *O*s were staff members of the laboratory and the rest were students in the University.

Electric shock.—The apparatus was essentially that used by Stevens, Carton, and Shickman (1958) to determine a ratio scale for the apparent intensity of electric shock. The *O* immersed the first and third fingers of his left hand (up to the first joint, approximately) in salt-water electrodes. A 60-cycle current was passed through these electrodes and through a series resistor of 250,000 ohms. The strength of the current through the fingers was determined by measuring the voltage drop across a 10,000-ohm resistor substituted for *O*'s fingers. The current was varied by means of a Variac transformer. A timer in the circuit limited the stimulus duration to .6 sec. at each presentation. The apparent intensities of five levels of current (.29, .41, .52, .62, and .72 milliamp.) were matched by having each of the 10 *O*s squeeze the dynamometer twice. Here, and in all the other experiments, the order of the presentation of the criterion stimuli was "random," except that neither of the two extreme values was ever presented first.

White light.—The target was a circular patch, of homogeneous luminance, subtending 5 degrees of arc when viewed from a distance of 17 in. The luminance of the patch was controlled by means of neutral density filters. The *O*, who was dark-adapted for about 10 min. before the session, looked through a binocular aperture at the end of a black hood enclosing the target. A key, pressed by *O*, operated a timer that lighted the source for 2 sec. There were six stimulus levels: 56, 66, 76, 86, 91, and 96 db re 10^{-10} Lambert. (In this notation, 70 db corresponds to 1 mL.). Each stimulus was matched twice by each of 10 *O*s.

White noise.—The output of a white noise generator was fed through an attenuator to a pair of PDR-8 earphones mounted in MX-41/AR cushions. The *O* turned on the stimulus by depressing a switch in the circuit. He was instructed to listen to the noise for 1–2 sec. at a time and to try to match the loudness by squeezing the dynamometer. Five levels were used as stimuli: 55, 65, 75, 85, and 95 db re .0002 dyne/cm.^{2}. Each was matched four times by each of 10 *O*s.

1000-cycle tone.—The apparatus was the same as that used in the experiment on white noise, except that the signal source was an audio oscillator set at 1000 cps. Six levels were used as stimuli: 47, 57, 67, 77, 82, and 87 db re .0002 dyne/cm.^{2}. Each was matched twice by each of 10 *O*s.

Vibration.—A 60-cycle signal from an audio oscillator was fed through an attenuator into a matching transformer whose output activated a V47 Goodmans Vibration Generator. The *O* rested his left middle finger tip lightly on a plastic button (about $\frac{1}{4}$ in. in diameter) affixed to the armature of the vibrator. He was instructed to turn on the stimulus by depressing a switch for a second or two at a time, and to try to match its intensity by squeezing the dynamometer. The six stimulus values were produced by impressing, across the terminals of the vibrator, voltages of 27, 37, 43, 49, 53, and 57 db re 1 mv. On this scale, threshold is in the vicinity of 10 db. Each stimulus was matched three times by each of 12 *O*s.

RESULTS

The median results for each of the five experiments are shown in Fig. 1, together with the results of the two earlier experiments (J. C. Stevens & Mack, 1959), in which *O*s endeavored to exert forces that matched the apparent heaviness of lifted weights and the apparent intensity of pressure applied to the passive skin of the palm. The points in Fig. 1 determine in log-log coordinates a set of functions that are approximately linear but widely different in slope. In other words, these equal-sensation functions approximate power func-

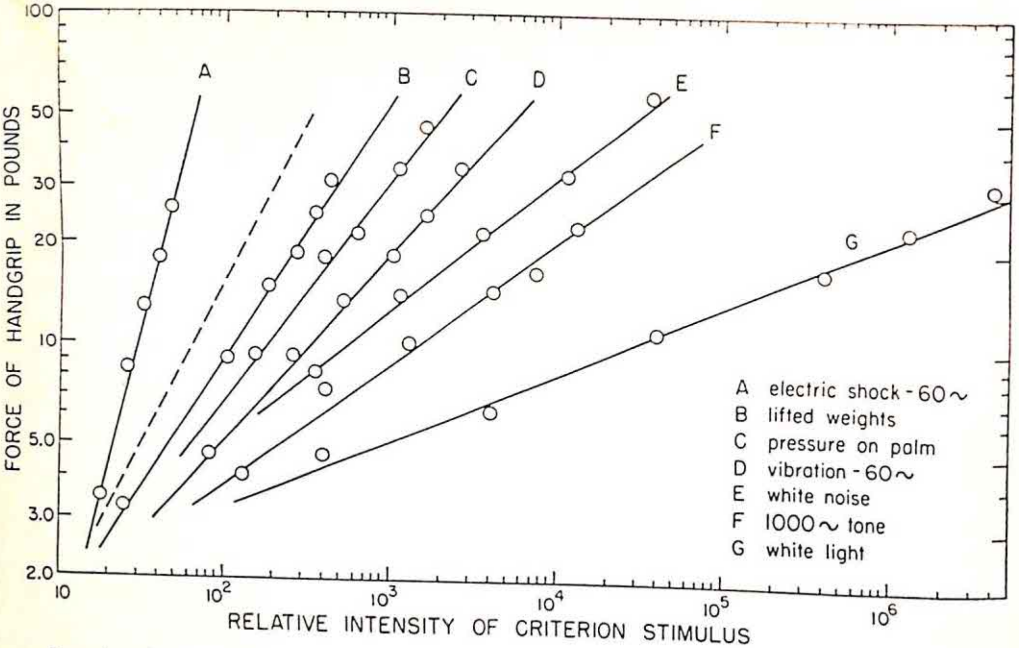


FIG. 1. Seven equal-sensation functions. The relative positioning of the functions along the abscissa is arbitrary. (For the absolute stimulus values, see text.) The dotted line shows a slope of 1.0 in these coordinates. Each point stands for the median force exerted to match a criterion stimulus.

tions having exponents of different sizes. The exponents (slopes) of the lines in Fig. 1 (determined by least squares) are listed in Table 1, where they may be compared with the exponents predicted from previous experiments in which the slopes were determined by "numerical" methods (e.g., magnitude estimation). The outcome of the handgrip experiments seems to confirm the form and the exponents of the several functions

TABLE 1

THE EXPONENTS (SLOPES) OF EQUAL-SENSATION FUNCTIONS, AS PREDICTED FROM RATIO SCALES OF SUBJECTIVE MAGNITUDE, AND AS OBTAINED BY MATCHING WITH FORCE OF HANDGRIP

Continuum	Ratio Scale		Exponents Determined by Handgrip	
	Exponent of Power Function	Reference	Predicted Exponent	Obtained Exponent
Electric shock (60-cycle current)	3.5	S. S. Stevens, Carton, & Shickman (1958)	2.06	2.13
Brightness of white light	0.33	S. S. Stevens & Galanter (1957)	0.20	0.21
Loudness of white noise (SPL)	0.6	J. C. Stevens (1958)	0.35	0.41
Loudness of 1000-cycle tone (SPL)	0.6	S. S. Stevens (1955, 1956)	0.35	0.35
60-cycle vibration	0.95	S. S. Stevens (1959b)	0.56	0.56
Heaviness of lifted weights	1.45	S. S. Stevens & Galanter (1957)	0.85	0.79
Pressure on palm	1.1	J. C. Stevens & Mack (1959)	0.65	0.67

determined independently by the ratio scaling methods based on numerical estimation.

Variability.—The variability of handgrip matching turns out to be about as great as the variability of numerical judgments of sensory magnitude. For a given continuum, the interquartile range of the forces exerted to match a criterion stimulus is approximated by the formula:

$$\log Q_3 - \log Q_1 = k$$

where k has a different value for each continuum matched. From Table 2, which lists the median value of k for each of the seven continua, it appears that it was harder to match some continua than others. The value of k for a 1000-cycle tone, for example, is approximately 2.5 times greater than its value for skin pressure. Differences of this kind were also reflected in the comments of some *O*s about the difficulty of the task. It is an interesting fact that the continuum that has been most assiduously explored in this country and elsewhere, namely, the loudness of a 1000-cycle tone, is perhaps the most variable and difficult of these seven. At any rate, the value of k seems to be roughly the same size as comparable measures of the variability of magnitude estimates (see S. S. Stevens, Carton, & Shickman, 1958; S. S. Stevens, 1956; J. C. Stevens, 1958).

TABLE 2

VARIABILITY OF THE MATCHES

k is the median of the interquartile ranges ($\log Q_3 - \log Q_1$). In order to obtain k', each *O*'s judgments (squeezes) were multiplied by a factor such that the judgments of all *O*s were reduced to a common modulus. k' is the median interquartile range ($\log Q_3 - \log Q_1$) of the values obtained by this procedure.

Continuum	k	k'
Electric shock	0.35	0.20
White light	0.30	0.22
White noise	0.26	0.20
1000-cycle tone	0.48	0.23
Vibration	0.36	0.18
Lifted weights	0.19	0.12
Pressure on palm	0.25	0.16

In the task of matching a continuum such as handgrip to each of various other continua, there are several sources of variability. In addition to instrumental errors and the inherent variability of human judgment, some of the variability is attributable to individual differences among *O*s. These latter differences seem to manifest themselves in two principal ways: as a difference in the slope and (or) as a difference in the position (intercept) of the equal-sensation function.

Although differences in slope may arise from several causes, they sometimes reflect genuine differences in the action of sensory transducers, e.g., recruitment in the ear or in the eye (S. S. Stevens, 1958). Actually, whether the slope is diagnostic of the operating characteristic of the sensory system depends greatly on what is matched to what. Thus differences in slope obtained with methods that involve numerical estimation usually arise from the simple fact of human variability. On the other hand, comparable differences in slope obtained by matching, say, loudness in one ear to loudness in the other, would generally indicate fundamental differences in the two loudness functions.

As regards the position of an equal-sensation function, differences here nearly always reflect differences in the judgmental criterion. That is to say, the absolute value of the force of handgrip exerted to produce a match may vary considerably from one *O* to another, for the simple reason that a "perfect" match is impossible and *O* must choose some criterion of what he will call equal. This effect is analogous to the use of a different modulus by each *O* in those magnitude-estimation experiments in which no standard stimulus is presented (S. S. Stevens, 1956). A major component of this variability can be removed by multiplying the results of each *O* by whatever factor is needed to bring the results into coincidence with those of the rest of the *O*s at some selected point.

This procedure of adjusting the judgments (squeezes) to a common "modulus," so to speak, was tried for each of the seven continua. The median results are shown in Table 2 under the heading k'. It is seen that the total variability is considerably reduced when the "modulus" factor is eliminated.

Inverse matchings.—The hand dynamometer lends itself reasonably well to the procedure under which *O* produces squeezes to match the apparent intensities of other sensations. It is adapted less well to the inverse task in which *O* adjusts the level of some other sensation to match the apparent force of a given squeeze. For this inverse procedure *O* is required to squeeze until *E* signals that the proper pull has been reached, and then *O* proceeds to adjust another stimulus (e.g., a sound intensity) until it seems to match the force he has exerted. Exploratory experiments were conducted with this procedure for two types of stimuli: white noise and 60-cycle vibration.

The results for the loudness of white noise were quite satisfactory, in the sense that the slope of the equal-sensation function was close to that shown by Line *E* in Fig. 1. As in other cross-modality matchings (S. S. Stevens, 1959a), there appeared a small "regression" or "centering" effect which made the function slightly steeper when the loudness was adjusted than when the force of handgrip was adjusted.

The results for vibration, however, were quite unsatisfactory, for the slope of the equal-sensation function was much steeper than expected. This large discrepancy is due mainly to the fact that *O*s do not have three hands. Rather than devise some novel method by which *O* could adjust the vibration, he was asked to squeeze the required amount with his right hand, and then to reach over and adjust, again with his right hand, the vibration amplitude applied to the middle finger of his left hand. This was apparently too much to ask, especially when matchings were required for a wide range of criterion forces. For six forces ranging from 4 to 37 lb., the function obtained had a slope of about 0.94 (compared with 0.56 for Line *D* in Fig. 1). This value dropped to about 0.73 when the range was made 5 to 25 lb. In a third experiment in which only three criterion forces (7, 12, and 20 lb.) were used, the slope was about .80. Needless to say, the variability of the vibration adjustments was large. In all three experiments the over-all range of the settings produced to match a given force was consistently between 20 and 30 db. It appears that *O* cannot "remember" a force well enough to carry out the delayed adjustment of a matching vibration with reasonable precision. Under these conditions the magnitude of the "regression" effect becomes extremely large.

Presumably a successful matching of vibration to handgrip might be possible if *O* could vary the amplitude of the vibration by means of some type of foot control. This would permit more or less simultaneous comparison of the two sensations.

DISCUSSION

One way of regarding the results presented in Fig. 1 is to see them as a possible reassurance to those who express misgivings concerning the use of numerical estimation in the measurement of sensory effects. In obtaining the results with the dynamometer, we have effectively avoided any demand on "number behavior," and we have not called upon any explicit concepts of ratio or proportion. All that was asked was that *O* try to assess the magnitude of a given sensation by producing another sensation of comparable intensity. The equating of the apparent strengths of two different sensations is a procedure sanctioned by much usage, as in the determination of equal loudness contours, heterochromatic brightness matches, etc., and presumably few people object to it on principle. In the present instance, the traditional procedure has been extended to the matching of apparent intensities across two modalities. Depending on the modalities involved, this task may be harder or easier than, for example, the matching of the loudnesses of two sounds having disparate qualities.

Granted that squeezes can be matched to other sensations, the hand dynamometer provides a method for measuring sensory intensity in the sense of determining the *relative* form of the over-all *operating characteristic* of the sensory system. Thus the data in Fig. 1 show that the sensation of electric shock grows with an acceleration about six times greater than loudness. In other words, the relative slopes of the functions in Fig. 1 have a significant meaning, for they describe the relative rates of growth of the various sensations. Furthermore, if we were willing to take any one of these sensory continua as a standard of reference, we could proceed to use it as a yardstick with which to measure sensory intensity on each of the other continua. This procedure would be useful for certain purposes, but the scales thereby achieved would have a basic limitation. Although they would presumably be power functions of the *linear* ratio scale of sensory intensity, the degree (exponent) of the power function would not be known.

What we would like, of course, would be to know the forms of the several functions relative to that of a linear continuum. But there seems to be in this world no *absolute* criterion of linearity; if everything in the universe were curved how would we know it? Hence the choice of a reference continuum is to some extent arbitrary. On the other hand, as a plausible reference continuum (the one assumed to be linear), the number continuum itself, when used by a person with proper understanding of it, has much to recommend it. The number system is readily accessible, and it "agrees" quite well with certain other continua (e.g., apparent visual length) on which linearity of subjective magnitude may reasonably be expected. What this means concretely is that, when *O*s are asked to assign numbers to apparent lengths of lines, the result is usually a power function with a slope close to 1.0. Thus a 4-in. line looks about twice as long as a 2-in. line.

On the assumption, then, that *O*'s number scale may be regarded as linear, it becomes a matter of interest to compare two kinds of experiments, one in which *O* emits squeezes to indicate magnitude of sensation and one in which he emits numbers for the same purpose. This comparison can be made in two stages. One question concerns the *relative* values of the exponents (slopes in log-log coordinates) obtained in the two experiments, and the other question concerns the *absolute* values of these exponents.

It is clear from Fig. 1 and Table 1 that the rank order of the slopes (exponents) is the same whether *O* matches the various sensory continua with squeezes or with numbers. Furthermore, to a good approximation, the slopes obtained by "handgrip matching" are a constant submultiple (1.7) of the slopes obtained by "number matching." To the extent that this proportionality obtains, we have a method by which the "relative" exponents obtained by squeezing can be converted into "absolute" exponents.

In the present instance the factor of proportionality (1.7) was derived from previous experiments designed to match, so to speak, number continuum against apparent force of handgrip. This kind of matching exhibits the usual "regression" effect referred to above, in the sense that when *O* adjusts handgrip to match numbers (magnitude production) the slope is steeper than when he adjusts numbers to match handgrip (magnitude estimation). This raises an interesting theoretical point. Which, if either, of these two slopes can be expected to correspond to the factor of proportionality? To the extent that both these slopes are biased by the "regression" effect, it seems that neither would correspond in general to the factor of proportionality. On the other hand, if it happens that the effects on the slope produced by interchanging standard and variable stimulus tend to be equal and opposite, then a slope intermediate between the two obtained slopes should provide the desired factor.

This, indeed, is what seems to have worked out. The value 1.7 for the exponent of the function relating number and handgrip was determined by pooling the results of magnitude production and magnitude estimation. With this relatively unbiased measure of the exponent determined, it becomes possible to substitute squeezes for numerical estimation, and then to recover the results characteristic of numerical estimation merely by multiplying the "squeeze" exponents by 1.7. The entries in Table 1 show how well this has worked out.

Among the advantages of validating the measurement of sensory intensity by methods that do not require *O* to use numbers, are the possibilities opened up for the diagnosis and study of idiosyncratic tendencies in numerical estimations. When a person's numerical estimates of a variable-like loudness produce a function that differs consistently and systematically from the median function, it is always an interesting question whether his hearing is different or whether he simply estimates differently. If his loudness function does not differ consistently from "normal" when determined by several cross-modality matchings, e.g., by squeezing, by matching loudness to vibration, etc., it would suggest that the dynamic operating characteristic of this person's auditory system is probably not peculiar. On the other hand, when the slope of the loudness function for a particular individual maintains a given peculiarity under several cross-modality matchings, evidence may be at hand that the peculiarity is real. In this fashion it would, in principle, be possible to diagnose such clinically important stigmata as recruitment, or its absence, in a hard-of-hearing person. Since the indicated therapy may be quite different in the two cases, this is one of those instances in which a valid measurement of the growth function governing a sensation is a matter of grave concern.

A final comment is in order regarding a theory that is alleged to account for the form of the various subjective scales. This theory revives an old notion about judgment, to the effect that estimations of sensory magnitude are based on experience with some physical measure of stimulus magnitude, i.e., the judge commits the "stimulus error." Warren (1958) elaborates on this idea and produces what he calls the *physical-correlate theory* of sensory intensity. This theory "considers that judgments of sensory magnitudes are based upon experience with the manner in which sensory excitation is correlated with the amount of some physical attribute associated with the stimulus" (Warren, 1958, p. 676). In his search for a plausible basis on which people might learn to judge such things as loudness and brightness, Warren comes up with the hypothesis that judgments of these continua are based on our experience with the inverse square law. (Actually the inverse square law does not predict correctly the exponents of the sone and bril scales, but Warren undertakes to explain away this defect by suggesting that the sone and bril functions were determined under "unfamiliar" conditions.)

A "theory" of this sort, since it is probably not susceptible to proof or disproof, must be judged, if at all, on the basis of the demands it makes upon our credulity. Some will find it difficult to believe that the relations among the exponents of all the functions in Fig. 1 are explainable on the basis of *O*'s experience with the physical stimuli involved. Where and how, for example, did *O*s "learn" the function for electric shock, or for vibration? And is it plausible to attribute to our prior experience with "physical correlates" the fact that the relations among the sensory scales remain invariant under a wide variety of cross-modality matchings?

Another objection to this theory is that it is essentially negative. If taken seriously, it would seem to imply that neither magnitude estimations nor cross-modality matchings are able to tell us anything about the operating characteristics of a sensory system. Pushed to an extreme, it would deny that we could tell the difference between a normal ear and an ear with recruitment, for if all sensory judgments are conditioned only by prior learning, the properties of the end organ and its associated neural structure cannot matter. Still, it is easy to change the form of a sensory function by altering the state of an end organ, for example, by light-adapting the eye.

Since the theory is consistent with the power-function relation between stimulus and response, it is claimed that "the physical-correlate theory provides a basis for Stevens' empirical law" (Warren, Sersen, & Pores, 1958, p. 707). Although it would indeed be nice to find a general basis for this psychophysical law, the law itself would seem to lose much of its interest if it turned out that the characteristics of the sensory transducers have nothing to do with the exponents of the various power functions. It is not, of course, unreasonable to suppose that experience with physical correlates is an important factor in determining numerical judgments on some kinds of perceptual continua (e.g., length of lines), and it is certainly true that an arbitrary scale can be learned (as when acoustical engineers learn to estimate sound levels in decibels). Nevertheless, it is not unreasonable to believe that the slopes of functions like those in Fig. 1 may depend more on the nature of the organism and its sensory transducers than on experience with such "physical correlates" as the inverse square law.

SUMMARY

The *O*s produced forces on a hand dynamometer to match various levels of subjective intensity for five kinds of criterion stimuli: electric current applied to the fingers, white light, white noise, a 1000-cycle tone, and a 60-cycle vibration applied to the finger tip. It is shown that the forms of the five equal-sensation functions (plus two functions obtained earlier for lifted weights and for pressure applied to the palm) can be predicted by the ratio scales of subjective intensity determined independently by the procedure of numerical estimation. The results add further evidence to the validity of the general psychophysical law that subjective magnitude grows as a power function of stimulus magnitude. Although force of handgrip cannot by itself be used to determine the absolute values of the power-function exponents, it can be used to determine the *relative* values of the exponents for a wide variety of prothetic continua. Taken in conjunction with the previously determined scale for the subjective magnitude of handgrip (exponent = 1.7), the matching of handgrip to various stimulus intensities confirms the approximate *absolute* values of the exponents for seven different continua.

REFERENCES

STEVENS, J. C. Stimulus spacing and the judgment of loudness. *J. exp. Psychol.*, 1958, **56**, 246–250.

STEVENS, J. C., & MACK, J. D. Scales of apparent force. *J. exp. Psychol.*, **58**, 405–413.

STEVENS, S. S. The measurement of loudness. *J. Acoust. Soc. Amer.*, 1955, **27**, 815–829.

STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

STEVENS, S. S. Some similarities between hearing and seeing. *Laryngoscope*, 1958, **68**, 508–527.

STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, 1959, **57**, 201–209. (a)

STEVENS, S. S. Tactile vibration: Dynamics of sensory intensity. *J. exp. Psychol.*, 1959, **57**, 210–218. (b)

STEVENS, S. S., CARTON, A. S., & SHICKMAN, G. M. A scale of apparent intensity of electric shock. *J. exp. Psychol.*, 1958, **56**, 328–334.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

WARREN, R. M. A basis for judgment of sensory intensity. *Amer. J. Psychol.*, 1958, **71**, 675–687.

WARREN, R. M., SERSEN, E. A., & PORES, E. B. A basis for loudness-judgments. *Amer. J. Psychol.*, 1958, **71**, 700–709.

(Received February 24, 1959)

Journal of Experimental Psychology
Vol. 59, No. 1, 1960

THE RELATIONSHIP BETWEEN DIRECTION AND AMOUNT OF STIMULUS CHANGE AND AMOUNT OF PERCEPTUAL DISPARITY RESPONSE[1]

H. D. KIMMEL

University of Florida

In a recent experiment (Grings & Kimmel, 1958) a difference in magnitude of GSR in favor of unexpected over expected stimuli was found on test trials following a training series in which *S*'s expectations had been formed by confirming experiences. This difference in magnitude of response was referred to as a "perceptual disparity response" (hereafter called PDR), the name having been adapted from the work of Bruner (1951). The training series in that experiment involved following *S*'s right or left lever movements by the presentation of a light or tone, the pairing of right-light, left-tone, right-tone, or left-light being maintained consistently throughout the training series. On test trials either the trained (expected) or the other (unexpected) stimulus was presented after a particular lever movement. The *S* was free to move the lever to either side, both during training and test, except that he was instructed to avoid marked favoring of one side over the other. The PDR was atattributed to a discrepancy between *S*'s expectation and what actually was presented.

In the present experiment the amount of discrepancy between expected and actual stimulation was varied along the quantitative stimulus dimension of physical intensity. The purpose of the study was to determine whether or not the amount of PDR is a function of the amount and/or the direction of the discrepancy between expected and actual stimulation. The stimulus dimension used to evaluate these questions was that of auditory intensity, defined in physical terms (i.e., decibels re .0002 dynes per cm^2). The data were gathered during the second part of an experiment on classical conditioning of the GSR, in which CS intensities had been varied along a physical continuum (Kimmel: 1958, 1959). Discrepancies between expected and actual stimulation were introduced at the beginning of an extinction series by changing the intensity that had been used during the acquisition series. It was assumed that *S* became set to receive the CS intensity experienced during acquisition, so that when the intensity was changed for the extinction series a discrepancy between expectation and actuality occurred. Sets and PDRs of this type have been demonstrated in previous research of this type (Grings, Dossett, & Honnard, 1959; Grings & Kimmel, 1958; Grings & O'Donnell, 1956).

Amount of PDR was specified for this study in terms of GSR magnitudes on the first trial after the change in stimulus intensity was introduced. The effect of stimulus intensity, per se, on amount of GSR (Hovland & Riesen, 1940) was controlled by correcting test response magnitudes for stimulus intensity differences in a manner described below.

---

[1] Based on data collected as part of a dissertation submitted to the Graduate School, University of Southern California, 1958.

METHOD

Subjects and design.—Ninety-six undergraduate students at the University of Southern California, 56 men and 40 women, served as volunteer *S*s.[2] Each *S* was paid $1.00 for participating in the experiment.[3] Volunteers with more than 15 db hearing loss in either ear were excused from participating. Two volunteers were rejected on this basis.

The *S*s were assigned randomly to 15 experimental groups according to the plan given in Table 1.[4] The groups differed according to the combination of acquisition and extinction stimulus intensities they received. As is indicated in Table 1, 13 different amounts of change in stimulus intensity were generated by this design, ranging from a −70 db change for the group shifted from an acquisition intensity of 115 db to an extinction intensity of 45 db, through zero change for the group receiving acquisition and extinction intensities of 75 db, to a +70 db change for the group shifted from an acquisition intensity of 35 db to an extinction intensity of 105 db. The groups shifted from acquisition intensities of 35 and 95 db to extinction intensities of 45 and 105, respectively, both experienced a +10 db change, while the groups shifted from acquisition intensities of 55 and 115 db to extinction intensities of 45 and 105 db, respectively, both experienced a −10 db change. Thus, the 15 combinations of acquisition and extinction intensities led to 13 amounts of change.

TABLE 1

NUMBERS OF *S*s AND AMOUNTS OF INTENSITY CHANGE (ΔI) ASSOCIATED WITH 15 COMBINATIONS OF CS AND EXTINCTION STIMULUS INTENSITIES (+ INDICATES INCREASE IN INTENSITY; − INDICATES DECREASE IN INTENSITY)

CS Intensity	Extinction Stimulus Intensity					
	45 db		75 db		105 db	
	N	ΔI	*N*	ΔI	*N*	ΔI
35 db	8	+10	8	+40	8	+70
55 db	4	−10	4	+20	4	+50
75 db	8	−30	8	0	8	+30
95 db	4	−50	4	−20	4	+10
115 db	8	−70	8	−40	8	−10

Apparatus.—The equipment used in this study has been described in detail elsewhere (Kimmel: 1958, 1959). A Hewlett-Packard Model 200-B audio oscillator was used to generate a 1,000 cps tone. The duration of the tone was 4 sec., controlled electronically. The tone was delivered to *S*'s PDR-8 headphones (with MX-41/AR cushions) through a Hewlett-Packard Model 350-B decibel attenuator. Intensity was rated at *S*'s headphones (in db re .0002 dynes/cm.^{2}) and was varied by changing the setting of the attenuator.

The GSR was picked up by a saline paste-metal-skin junction using $\frac{3}{4}$-in. Lucite cups attached by rubber straps to the palm and back of *S*'s nonpreferred hand. Alcohol was used to clean the skin before the electrodes were attached. The response was measured as a resistance change in a Wheatstone bridge input to a Brush DC amplifier, with changes in resistance recorded as deflections on an Esterline-Angus ink-writing milliammeter with a paper speed of 3 in. per min. The current through *S* was constant at 42.5 microamperes.

Instructions.—Instructions were administered in writing. They informed *S* that this was an experiment on the way in which different stimuli evoke the GSR. No task was given to *S* other than attending to the stimuli, remaining as motionless as possible, and relaxing his body to prevent discomfort.

Procedure.—Observations were made in a soundproof room with all lights out except for a covered 7.5-w. lamp in *E*'s cubicle. A partition separated *S* from *E* and the equipment. The only noise in the room was the click of the signal magnets on the recorder and the hum of an electric fan masking these clicks.

---

[2] Sixteen *S*s were run by R. Honnard.

[3] Payments to *S*s came from a grant from the National Science Foundation to W. W. Grings.

[4] The purpose of the first part of the study was to evaluate the effect of CS intensity on amount of conditioning during acquisition and extinction. A 5 × 3 design, replicated four times for *S*s receiving paired tones and shocks, was generated with this purpose in mind. A 3 × 3 design, replicated four times for *S*s receiving unpaired tones and shocks, was added to provide controls for sensitization effects in the paired group. Since the study reported here was a secondary consideration at the time of data collection, the *N*s in each of the 15 groups are not the same, nor are they the same in the 13 groups receiving different amounts of intensity change.

After receiving a hearing test and reading the instructions, *S* was blindfolded and the equipment attached. The *S* then received three presentations of the CS intensity and three of the extinction stimulus intensity (both being 1,000 cps tones of 4 sec. duration) in a counterbalanced order fixed for all *S*s. This preliminary presentation series was followed by three shocks to *S*'s preferred forearm (the shock intensity was set by *S* at the beginning of the session so that it would be mildly annoying) and six additional presentations of the CS intensity. These trials were all separated by 20–30 sec., varied unsystematically. Sixty *S*s then received 20 trials, separated by 40–60 sec., of the CS paired with shock (these 60 *S*s were divided into 15 groups of 4, each group receiving a different one of the 15 combinations of acquisition and extinction stimulus intensities), while the remaining 36 *S*s received a random sequence of unpaired presentations of 20 shocks and 20 tones, separated by 20–30 sec., varied unsystematically (these 36 *S*s were divided into 9 subgroups of 4, each group receiving a different one of the 9 combinations of acquisition and extinction intensities excluding those involving the 55 and 95 db acquisition intensities).

All 96 *S*s then received six presentations of the extinction stimulus without shock. The first of these extinction presentations provided the dependent measures for this report.

RESULTS

Magnitude of GSR was measured as the maximum change in skin resistance occurring within a 5-sec. interval beginning with the onset of the test stimulus. This measure was transformed into units of square root of conductance change. The transformed responses obtained on the three preliminary presentations (i.e., before acquisition) of the extinction stimulus were averaged separately for each *S* and this average was subtracted from *S*'s transformed response on the first extinction trial to provide a measure of PDR corrected for the effect of stimulus intensity differences, and, to some extent, for differences among *S*s in GSR responsiveness.

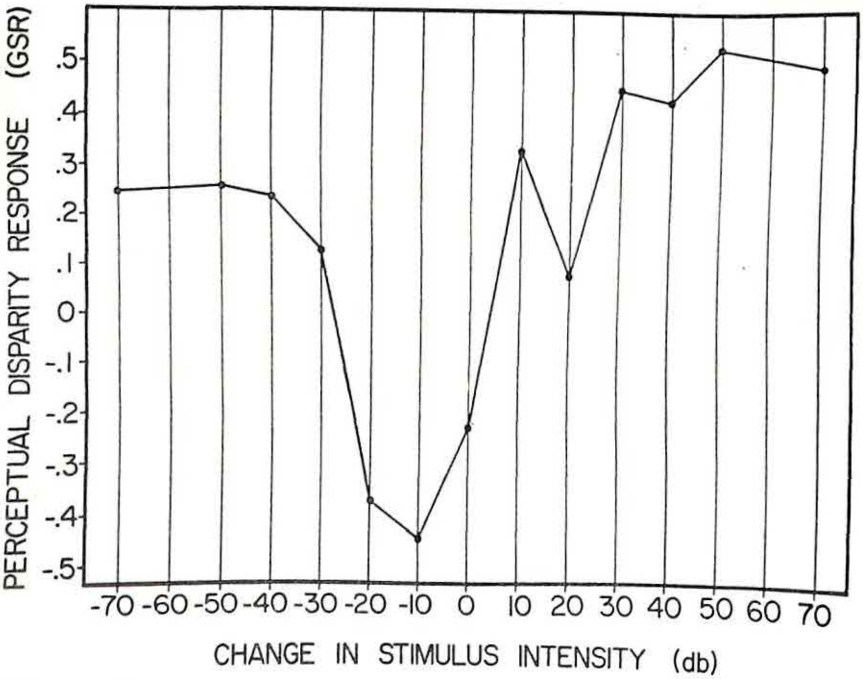


FIG. 1. Amount of perceptual disparity response as a function of amount of change in stimulus intensity.

Figure 1 is a graphical representation of the differences among mean PDRs associated with the 13 amounts of stimulus intensity change. The figure indicates that greater mean PDRs resulted from large amounts of intensity change than from small amounts of intensity change, regardless of the direction of the intensity change, and that greater mean PDRs resulted from increases in intensity than from decreases in intensity.

To simplify the statistical analysis of these data, *S*s were divided into four subgroups according to whether they had received positive or negative intensity changes and whether they had received small or large amounts of intensity change. Thus 24 *S*s who had received positive changes of 10, 20, or 30 db constituted one subgroup, 24 *S*s who had received negative changes of 10, 20, or 30 db formed a second subgroup, 20 *S*s who had received positive changes of 40, 50, or 70 db a third subgroup, and 20 *S*s who had received negative changes of 40, 50, or 70 db a fourth subgroup. The 8 *S*s who received zero change in intensity were omitted. Table 2 presents the means and *SD*s of PDRs for these four subgroups and illustrates the logic of the analysis of variance of PDRs presented in Table 3.

TABLE 2

PDR MEANS AND *SD*s ON THE FIRST EXTINCTION TRIAL

Amount of Change in Intensity		Direction of Change in Intensity		Totals
		Upward	Downward	
Small (10, 20, 30 db)	*M*	0.32	−0.24	0.04
	SD	2.44	1.89	
	N	24	24	
Large (40, 50, 70 db)	*M*	0.47	0.24	0.36
	SD	2.66	2.36	
	N	20	20	
Totals	*M*	0.39	−0.02	0.18

TABLE 3

ANALYSIS OF VARIANCE OF PDRs ON THE FIRST EXTINCTION TRIAL

Source	*df*	*MS*	*F*
Amount of intensity change	1	2.146	5.38*
Direction of intensity change	1	3.752	9.40**
Amount × Direction	1	0.623	
Error (within cells)	84	0.399	

* $P = .05$.
** $P = .01$.

As is shown in Table 2 and evaluated statistically in Table 3, the mean PDR for large changes in intensity was significantly greater than the mean PDR for small changes in intensity ($F = 5.38$, 1 and 84 *df*) and the mean PDR for changes in an upward direction was very significantly greater than the mean PDR for changes in a downward direction ($F = 9.40$, 1 and 84 *df*). The absence of an interaction effect indicates that these differences occurred at both levels of both variables.

DISCUSSION

The results of this study show quite clearly that the PDR observed by Grings and Kimmel (1958) is, at least in part, a function of the amount and direction of disparity between *S*'s expectation and actual stimulation, when stimulus intensity is the dimension along which the disparity occurs. In the present study a minor source of confounding was associated with the fact that the 35-db CS intensity was the only acquisition intensity that led to significant conditioning during the acquisition series (Kimmel: 1958, 1959). Since 12 *S*s who had been conditioned with the 35-db CS all experienced increases in intensity, it is possible that the significant difference in mean PDR favoring upward over downward changes partly resulted from generalization of this conditioning effect. Two considerations argue against this possibility, however, The first is the fact that 12 other *S*s who had received the 35-db CS had not paired with shock and thus could not have been conditioned. Secondly, the differences in amount of conditioning that had been observed during the acquisition series had all but vanished before the end of the acquisition series and, thus, could not have carried over to the first extinction trial.

It is not surprising that PDR, or as it might be referred to in more general terminology, startle, was found to vary positively with the amount of difference between what *S* expects to receive and what he actually does receive. If startle results from receiving the unexpected, more startle should be found when the difference between expectation and actuality is great than when it is small. Of considerable significance, however, is the implication of the present findings that *S*'s expectation is sufficiently stimulus-bound, in the physical sense of the word "stimulus," that it is sensitive to differences in both amount and direction of intensity differences.

It would be of interest to determine whether the effects observed in this experiment can be duplicated using stimulus dimensions other than intensity and sense modes other than audition. Earlier studies (Grings, Dossett, & Honnard, 1959; Grings & Kimmel, 1958) found the effect when changes were made between sense modes and were, therefore, more qualitative than in the present study.

SUMMARY

A study was done, using 96 *S*s, in which the intensity of an auditory stimulus (a 1,000-cps tone of 4 sec. duration) used during a conditioning series of 20 trials was changed for an extinction series. Thirteen different amounts of change in intensity were used, ranging from a 70 db downward change, through zero change, to a 70 db upward change. Amount of GSR, corrected for the effect of stimulus intensity differences, was used to measure the perceptual disparity response (PDR) occurring on the first trial with the changed intensity.

The results showed that amount of PDR was greater for intensity changes in an upward direction than for those in a downward direction and was greater for large intensity changes than for small intensity changes. Both of these effects were statistically significant, but there was no interaction between them. The results were interpreted to indicate that *S*s develop preparatory sets, or expectations, regarding the stimulation they will receive and based on the stimulation they have already received. These expectations appeared to be closely tied to the physical stimulus since they were highly sensitive to changes in physical intensity.

REFERENCES

BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey, *Perception: An approach to personality.* New York: Ronald Press, 1951.

GRINGS, W. W., DOSSETT, W. F., & HONNARD, R. Conditioned GSR in perceptual disparity situations. *Amer. Psychologist*, 1959, **14**, 393. (Abstract)

GRINGS, W. W., & KIMMEL, H. D. Magnitude of GSR to expected and unexpected stimuli. Paper read at Western Psychol. Ass., Monterey, April, 1958.

GRINGS, W. W., & O'DONNELL, D. E. Magnitude of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1956, **52**, 354–359.

KIMMEL, H. D. Amount of conditioning and intensity of conditioned stimulus. Unpublished doctoral dissertation, Univer. of Southern California, 1958.

KIMMEL, H. D. Amount of conditioning and intensity of conditioned stimulus. *J. exp. Psychol.*, 1959, **58**, 283–288.

(Received February 24, 1959)

Journal of
Experimental Psychology

VOL. 59, NO. 2 FEBRUARY 1960

EFFECT OF FIRST-ORDER CONDITIONAL PROBABILITY IN A TWO-CHOICE LEARNING SITUATION[1]

NORMAN H. ANDERSON[2]

University of Wisconsin

The experiments reported here investigate behavior in a two-choice probabilistic task in which the independent variable is the conditional probability in the sequence of stimulus events. Each of the two stimulus events (lights) was required to occur equally often, but a partial patterning was introduced by varying the first-order conditional probability in the sequence of lights. This variable, denoted by π_{11}, is defined as the probability of occurrence of Light 1 on any trial, given that Light 1 occurred the previous trial. When $\pi_{11} < .5$, the lights tend to alternate on successive trials; when $\pi_{11} > .5$, each light tends to repeat itself. Thus, $\pi_{11} = .7$ represents a 70% repetition pattern, and $\pi_{11} = .3$ represents an equally strong alternation pattern. Since each light occurs equally often, the appropriate dependent variable is the frequency of *repetition responses*, a repetition response being defined as the response of predicting on the current trial that same *light* which occurred on the previous trial. An alternation response is analogously defined as predicting that light which did not occur on the previous trial.

This conditional probability variable was first used by Hake and Hyman (1953). It is of considerable interest because of its close relationship to the *frequency* variable for which the Estes and Straughan (1954) model predicts the "matching" behavior first reported by Grant, Hake, and Hornseth (1951). Engler (1958) tested a Bush-Mosteller (1955) model which predicts both the absolute frequencies of the two choices, and the level of repetition responses. The model succeeded with the first dependent variable, but failed to fit the repetition response results.

The first of the experiments reported here was designed to test a model (Anderson, 1956) for the conditional probability situation. The model is based on the ideas of Estes (1950) and of Restle (1955). It is assumed that there are two distinct subsets of stimulus elements, one of which is specific to the repetition response, the other specific to the alternation response. On any trial the sampled stimulus elements specific to a given response become conditioned to that response if it was indicated as correct, regardless of whether *S* actually made the response. However, if the given response was not the correct one, then the sampled stimulus elements specific to that response go into a neutral state. The probability of a repetition response on any trial is the ratio of the number of sample elements conditioned to the repetition response to the total number of sample elements in the conditioned state. Neutral elements have no effect on response probability, but they may be conditioned on later trials. (The model also postulated a third set of stimulus elements corresponding to responses other than the two above. It was assumed that such stimulus elements were progressively neutralized over trials. This third set of elements would thus affect the learning rate, but not the asymptotic level. Hence, it need not be further considered here.)

---

[1] Based on work submitted in partial fulfillment of the requirements for the PhD degreee at the University of Wisconsin, 1956. The work was done while the author was an NSF predoctoral fellow. I wish to express my thanks to D. A. Grant, W. J. Brogden, and H. F. Harlow for their advice and helpful criticism. Thanks are also due S. H. Sternberg and R. C. Atkinson for a critical reading of the manuscript.

[2] Now at the University of California, Los Angeles.

This model leads to the following approximate expression for the asymptotic probability of repetition responses:

$$\frac{1}{1 + c(1 - \pi_{11})/\pi_{11}} \qquad [1]$$

where π_{11} is the conditional probability of Light 1, given that Light 1 occurred on the previous trial, and c is a constant which measures the relative weights of the two sets of stimulus elements.

Since the completion of the present work, Burke and Estes (1957) have revised the original Estes and Straughan (1954) model, taking into account the nonindependence of the sample of active stimulus elements on successive trials. The revision, it may be noted, yields some small changes in the predicted asymptotes for the event frequency situation. More pertinently, it also yields the following expression for asymptotic repetition response level:

$$\pi_{11} + \theta\pi_{11}(1 - \pi_{11}) \qquad [2]$$

where θ is the learning rate. This model will also be tested against the present data.

Upon completion of Exp. I, it was seen that the model, although it gave a reasonable account of the asymptotic acquisition performance, was unable to handle the behavior in the transfer condition in which the various acquisition groups failed to converge to a common level despite the fact that they were receiving the same treatment. Since this differential transfer effect appeared to create difficulties for the mathematical approach in general, it seemed desirable to abandon the testing of specific models and begin a more extensive empirical investigation. Experiment II was accordingly designed to give a parametric treatment of the conditional probability variable. The data were recorded so as to permit the easiest tabulation of the sequential dependencies in the stimulus-response sequences. It was expected, both on theoretical grounds (Anderson, 1959) and from the results of earlier work (Estes & Straughan, 1954; Anderson & Grant: 1957, 1958; Jarvik, 1951), that these dependencies would give additional insight into the processes underlying the behavior in this situation.

METHOD

The procedure common to the two experiments will be given first; that particular to the individual experiments will be found under the corresponding headings. Each group is labeled by 100 times its acquisition value of π_{11}. Thus, Group 30 received $\pi_{11} = .3$ in acquisition, etc.

Apparatus

The apparatus was the same as that used in previous work (Anderson & Grant, 1957). Two 6-w. opal lights, mounted 10 in. apart horizontally on a vertical, flat black, 15 × 20-in. panel, formed a stimulus display. The two lights were activated by a punched tape fed through a Western Union tape transmitter. An 800-cps tone served as the signal to respond. Presentation of tone and lights was automatic.

A row of four partially separated booths was placed 9 ft. from the panel of lights. Each booth contained a small box with two non-self-releasing keys, corresponding to the two lights. The *S*s were required to push the keys down, and then push them up, in making their responses. Stimuli and responses were recorded automatically with an Esterline-Angus operation recorder.

Except for the panel of lights and the booths, all equipment was placed on *E*'s desk which lay behind the booths, and from which *E* monitored the experiment.

Procedure

The *S*s were volunteers from elementary psychology classes who received class points for their work. They were randomly assigned, in groups of two, three, or four, to the various experimental conditions.

At the beginning of the experiment, *S*s were seated in the booths and read the instructions which were similar to Instructions C of Anderson and Grant (1957). In addition to operating detail concerning the keys, tone, and lights, the instructions stated that the task was to predict which of the two panel lights would come on next, that this was an experiment in learning, and that *S*s should do the best they could even though the task seemed difficult. Three practice trials, in which both lights flashed each time, were then given, and questions answered. Finally, *S*s were told that the rest of the experiment would have to be run off without conversation or interruption. The sequence of trials was continuous, with no external indication of change from acquisition to transfer condition.

Independent Variables

The two experimental variables were trial rate, and first-order conditional probability. The temporal division of each trial was as follows: signal tone (during which responses were to be made), 2.0 sec.; waiting time, 1.2 sec.; stimulus light on, .8 sec.; "intertrial interval," 1.0, 3.5, or 6.0 sec. These intertrial intervals thus yielded trial rates of 5.0, 7.5, and 10.0 sec. per trial.

First-order conditional probability has been defined in the introduction. The sequences of lights were random subject to the following rules: (*a*) exactly one light flash each trial; (*b*) the assigned value of π_{11} hold exactly within each block of 50 trials; (*c*) the absolute frequency of each light lie between 20 and 30 in each block of 50 trials. Four independent sequences were constructed for each acquisition value of π_{11}. In the transfer condition, four independent sequences were also used, but these were the same for all groups within each experiment.

Experiment I.—This experiment was based on a 3 × 2 design using the three trial rates, 5, 7.5, and 10 sec. per trial, and the two acquisition values of conditional probability, $\pi_{11} = .3$, and $\pi_{11} = .7$. Light sequences were kept orthogonal to these two variables. Each group received 200 acquisition trials. In the transfer condition, $\pi_{11} = .5$ for all groups. The 5, 7.5, and 10 sec. groups were given 300, 200 and 100 transfer trials, respectively.

Experiment II.—This experiment was a parametric study of the effect of conditional probability on acquisition, on transfer to an extinction condition, and on recovery. A group of *S*s was run at each of the 11 acquisition values of π_{11} ranging from 0 to 1 in steps of .1. The transfer condition was the same for all groups, with $\pi_{11} = .5$ as in Exp. I. Acquisition and transfer were given in Session 1; recovery was tested in Session 2.

Trials were presented at the 5-sec. rate used in Exp. I. New stimulus sequences were constructed for conditions duplicating those of Exp. I. Instructions were reworded slightly by incorporating an additional statement to the effect that, while *S* should be able to do considerably better than just guessing, a perfect score was impossible, so that *S* should not worry if he was not correct every time.

Groups 10–90 were given 300 acquisition trials, followed by 200 transfer trials. Groups

0 and 100 received 100 acquisition trials, and 400 transfer trials. For these two groups, right and left were counterbalanced in Trials 1–100 within each of the four transfer sequences. Twenty *S*s were run in each group, except that Group 50 contained 40 *S*s.

Two additional subgroups, each with $N = 12$, were run with $\pi_{11} = .3$, and .7. The sequences for these groups were identical with those of the corresponding conditions of Exp. I for the first 200 trials, and with those of Exp. II for the remaining trials. These two groups permitted an evaluation of effect of the change in instructions from Exp. I to Exp. II, without confounding any possible effect of sequences. Since no effect of instructions was found, these *S*s have been combined with the others so that Groups 30 and 70 each have $N = 32$.

Recovery was tested by giving 60 additional trials with $\pi_{11} = .5$. Fifteen such sequences were constructed by taking successive blocks of 60 trials from three of the transfer sequences of Exp I. The *S*s were assigned to these sequences unsystematically, but each sequence was used about equally often. Session 2 was run after an interval of two to six days. An *S* was signed up for Session 2 before serving in Session 1, according to *E*'s schedule and each *S*'s own convenience. In contrast to Session 1, *S*s were occasionally run alone in Session 2.

One *S* adopted the deliberate policy of leaving his key down for several trials until he wished to change his choice of responses. He was replaced although his record seemed normal in other respects. Three *S*s, all from the additional subgroup run under the .7 condition, did not return for Session 2. The Session 2 record of one *S* was lost because of a recording error.

RESULTS

A repetition response is defined as a press of that key corresponding to the light which flashed the previous trial. Since two stimuli and two responses are involved, a nonrepetition response may be called an alternation response. For conditions with $\pi_{11} > .5$, a repetition response is also an optimal response since it is more likely to be correct. Similarly, when $\pi_{11} < .5$, an alternation response is an optimal response. The main dependent variable is repetition response frequency.[3]

Unless otherwise noted, the analysis of variance was used in the statistical analyses. The Sequence variable, corresponding to the different random sequences used, was treated throughout as a fixed variate (Wilk & Kempthorne, 1955, p. 1163).

Mean Performance Curves

Experiment I.—The results of Exp. I are shown in Fig. 1, which plots mean percentage of repetition responses as a function of trial blocks. Acquisition is rapid, but marked reversals occur in the third block of 10 trials for Groups 30 (alternation groups), and in the fifth block of 10 trials for Groups 70 (repetition groups). Terminal response levels, calculated over Trials 151–200, average 79% for Groups 70, and 33% for Groups 30. These values correspond to optimal response levels of 79% and 67%, respectively. No trial rate effect is apparent for Groups 70, but the slower rate conditions of Groups 30 do somewhat poorer in the middle 100 acquisition trials. There is an immediate change in the transfer trials, in which all groups received the same purely random sequences. However, Groups 30 and 70 show little sign of approaching a common final level, even after the 300 transfer trials given the two 5-sec. groups.

Analyses of acquisition performance on total responses over Trials 1–200,

---

[3] Between .003 and .004 of the responses were "improper," with *S* pressing too late, pressing neither key, or, most frequently, pressing both keys. Frequency of such improper responses did not appear to depend on trial rate, conditional probability, or trial number, except that Groups 0 and 100 made no improper responses in acquisition. All improper responses were filled in randomly in order to simplify the tabulations.

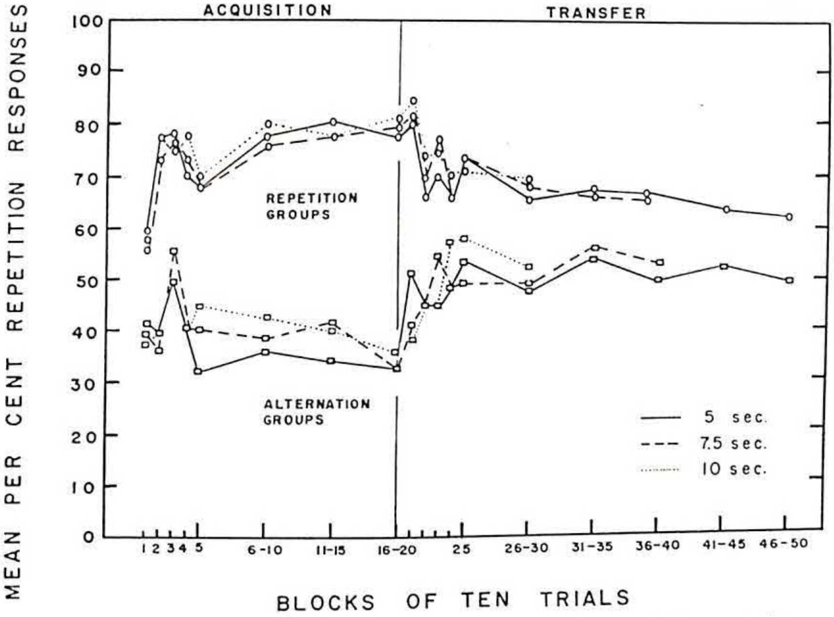


FIG. 1. Mean percentage of repetition responses in acquisition and in transfer as a function of trial blocks, Exp. I.

and over Trials 151–200 were performed for Groups 30 and Groups 70 separately. There were no significant results in the analyses of Groups 70, although the Sequence effect approached the .05 level in the first analysis. For Groups 30, the Sequence effect was significant in both analyses: *F* (*df* = 3, 48) = 5.46, and 4.40, for the analyses over Trials 1–200, and Trials 151–200, respectively. No other reliable differences among Groups 30 were found, although the linear component of trial rate approached significance, *F* (*df* = 1, 48) = 3.82 in the analysis of the total score over Trials 1–200. The comparison of optimal response frequencies, based on the pooled error term, yielded significant differences between the two conditional probability conditions: *F* (*df* = 1, 96) = 97.7 and 51.1, for the tests over Trials 1–200 and Trials 151–200, respectively.

Analyses of transfer performance were made on repetition response frequencies, and included Sequences as an additional factor orthogonal to the two main independent variables. The effect of the conditional probability factor was significant in all analyses. In particular, total repetition responses over Trials 451–500 for the two 5-sec. groups were significantly different, *F* (*df* = 1, 32) = 5.15. No other main effect or interaction approached significance.

The first 50 trials of the sequences of reinforcing lights were inspected for peculiarities which might explain the reversals early in acquisition. The reversal for Groups 70 apparently arose entirely from a run of seven consecutive alternations which occurred in one of the sequences. Nothing unusual was seen in the sequences for Groups 30, however. See also below.

Experiment II.—Figure 2 plots the mean percentage of repetition responses as a function of trial blocks with acquisition value of π_{11} as a parameter. Following the rapid acquisition, the various groups level off in the expected order with some conditions showing further slow changes. Three conditions deserve special mention. Group 30 exhibits an anomalous reversal similar to that seen in Exp. I. Group 40 does well in terms of optimal responses on Trials 1–10, but then ascends to the "chance" level where it remains for the duration of Session 1. Group 50 stabilizes at a repetition response level of 62% although the sequences of lights were entirely random in this condition. This tendency toward repetition responses in Group 50 is consistent with the finding of Exp. I that Groups 70 made more optimal responses than Groups 30. This latter result appears again in Exp. II, and the same trend is shown in the comparisons of Groups 40 and 60, and of Groups 20 and 80.

Table 1 gives the means and *SE*s obtained from total repetition response frequencies over Trials 201–300. This table includes the theoretical asymptotic response levels predicted by the author's model, and by the Burke and Estes (1957) model. The single unknown parameter in each model was evaluated from the data for Group 50 by equating obtained and theoretical asymptotes for this condition. This procedure yielded $c = .618$ for the present model, and $\theta = .472$ for that of Burke and Estes. The two models make essentially the same predictions, and these predictions are quite good for the larger values of π_{11}. It is clear, however, that neither model gives an adequate account at the lower values of π_{11}, where the groups show an increasing tendency to lie below the theoretical asymptote.

TABLE 1

OBTAINED AND PREDICTED MEAN PROPORTION OF REPETITION RESPONSES ON TRIALS 201–300, EXP. II

Group	Obtained	Predicted		s_m
		Eq. 1	Eq. 2	
90	.94	.94	94	.011
80	.90	.87	.88	.014
70	.84	.79	.80	.014
	(.77) (.79) (.80)			
60	.75	.71	.71	.020
50	.62	.62	.62	.021
40	.49	.52	.51	.025
30	.35	.41	.40	.014
	(.33) (.33) (.35)			
20	.16	.29	.28	.020
10	.04	.15	.14	.007

Note.—Entries in parentheses are from Trials 151–200 of Exp. I.

As is evident in Fig. 2, Groups 0 and 100 were shifted to the transfer condition after only 100 acquisition trials. The main results for these two conditions will be given in this paragraph and they will not be considered again unless explicitly mentioned. Mean errors in acquisition were 1.35 and 2.95 for Groups 0 and 100, respectively. The difference is significant as shown by the Mann-Whitney *U* test, $z = 2.93$. The groups overshoot each other in the fourth block of transfer trials, and Group 0 makes significantly more repetition responses in that block, F $(df = 1, 32) = 6.59$ by an ad hoc test. The response rate over Trials 201–500 was quite stable, but these data are omitted in Fig. 2. Group 0 averaged 3.7% fewer repetition responses than Group 50, and Group 100 averaged 2.3% more repetition responses than Group 50 over these last 300 trials. However, the between-groups F for these data was not significant.

In the transfer trials, all conditions show some initial change in response rate except for Groups 40, 50, and

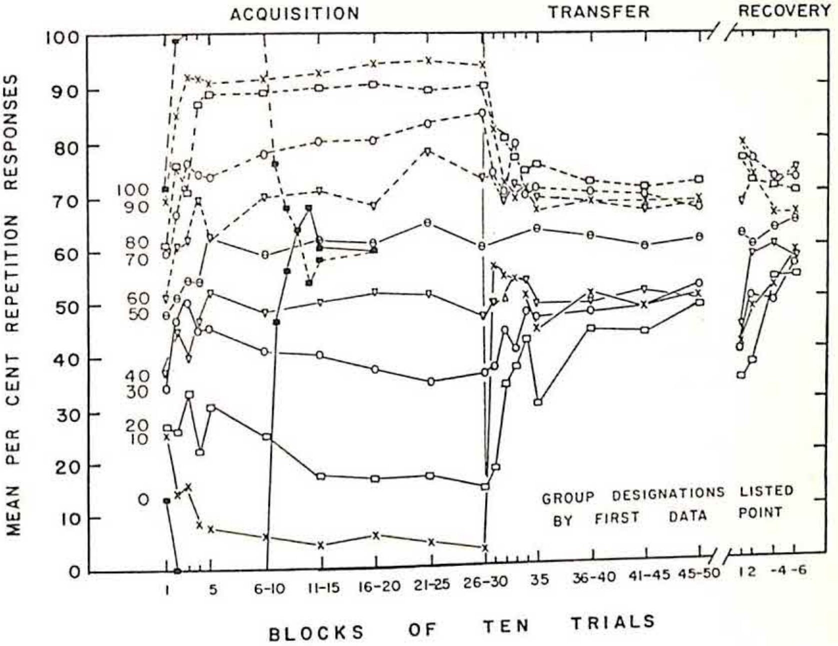


FIG. 2. Mean percentage of repetition responses in acquisition, in transfer, and in recovery, as a function of trial blocks, Exp. II.

60. Group 10 exhibits a sharp increase comparable to Group 0. The terminal transfer levels appear to be undergoing little change. However, these terminal levels are not smoothly related to acquisition treatment: Groups 60–90 all cluster around the 70% level, Groups 10–40 level off at about 50%, and Group 50 falls neatly between.

An analysis of total repetition responses over Trials 401–500 was performed with sequences of reinforcing lights and acquisition treatment as orthogonal factors. For acquisition treatment, $F\ (df = 8,\ \ 188) = 16.8$, but no other effects were significant. Using the value $s_m = 2.64$ (for a group of 20 *S*s) obtained from this analysis in Duncan's (1955) range test, Groups 10–40, Group 50, and Groups 60–90 separated out as three significantly different subsets of conditions, with no further differentiation within these three subsets.

Recovery is evident for most groups in Fig. 2, but dissipates in about 20 trials. Although Session 2 was too short to be more than suggestive, there seems to be a tendency for Groups 60–90 to return to their final level in Session 1, and for Groups 10–40 to rise above their final level in Session 1.

Recovery scores were computed by taking the absolute value of the number of repetition responses on Trials 481–500 from twice the number of repetition responses on Trials 1–10 of Session 2. A positive score thus indicates a recovery of the acquisition response. Groups 0, 50, and 100 were tested together but the analysis yielded no significant results. The mean recovery score, averaged over the remaining eight groups, was significantly different from zero, $F\ (df = 1,\ 173) = 32.3$. Further tests are vitiated by the lack of control over the interval between sessions.

Except for Group 100, right and left were not counterbalanced in the light sequences because position preferences cannot produce artifacts with the present independent variable. Position preferences may exist, of course, and two tests were made for them. First, all key presses for all *S*s over Trials 451–500 were pooled. There was a slight excess of left key presses, but a binominal test failed of significance. Second, all repetition responses of Group 50 over Trials 1–50, and over Trials 1–500 were treated similarly. No significant positional effects were found in these two tests either.

Sequential Dependencies

Considerable additional information may be extracted from the data by considering the dependence of the response on any trial upon the subsequence of stimuli and responses on the trials immediately preceding. The analysis of these sequential dependencies has proved useful in previous work on specific models (Anderson & Grant: 1957, 1958; Estes & Straughan, 1954). Without a specific model, however, the interpretation of the sequential dependencies presents certain pitfalls which have given rise to some confusion in the literature. Accordingly, some comments will be made at appropriate points on limitations in the interpretation of such data.

The two stimulus events will be denoted by 0 and 1. A subsequence of consecutive events will be called a *tuple*. The symbol, $R'(T)$, will denote the proportion of repetition responses conditional on the occurrence of any particular tuple, T. Thus, $R'(100)$ denotes the proportion of repetition responses, given that the stimulus lights on the three preceding trials were 1, followed by 0, followed by 0. Its value was calculated by dividing the number of occurrences of the stimulus tuple, 100, into the number of times that this tuple is followed by a repetition response which, in this case, would be the response of predicting Event 0. Because right and left were equiprobable in all light sequences, and because no position preferences were found, the data from each pair of positionally complementary tuples were pooled before dividing to obtain the conditional proportions. Consequently, $R'(100)$ and $R'(011)$ denote the same quantity, and this quantity is obtained by pooling the data from the two 3-tuples, 100, 011.

The specific tabulational procedure used here requires comment in two respects. In the first place, the compilations were done by pooling the data over all *S*s and all sequences in each condition before computing the desired proportions. There can result, as a consequence, certain biases as discussed more fully below. Second, the method of tabulation was such that, for instance, the tuple, 0000, contributed one instance to $R'(0000)$, two instances to $R'(000)$, three instances to $R'(00)$, and four instances to the repetition response rate, $R'(0)$. Similarly, the alternation tuple, 1010, contributed one instance to $R'(1010)$, two instances to $R'(010)$, three instances to $R'(10)$, and four instances to the repetition response rate, $R'(0)$; (see penultimate sentence of preceding paragraph). This procedure averages the R' values over all possible antecedents of the corresponding tuples, and thus places the R' values for runs, alternations, and other tuples on an equal footing. A somewhat different tabulational procedure has been used by Nicks (1959) in that part of his report dealing with run curves. The present method leads to simpler mathematical expressions relating sequential dependencies and model parameters. However, Nicks' method may prove to be more useful for nonparametric analyses.

It may be worth mentioning that these tabulations, especially those involving response dependencies, are time consuming. Consequently it is advisable to plan the general tabulation procedures and the data recording sheets with care so as to maximize efficiency. The use of the symbols 0 and 1, rather than

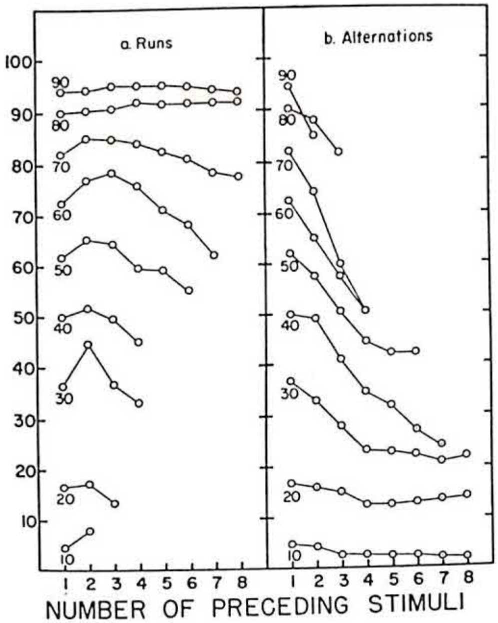


FIG. 3. Mean percentage of repetition responses over Trials 101–300, inclusive: (*a*) response following runs of the designated length; (*b*) response following alternation tuples of the designated length. See text.

R and L, is recommended since they are easier to write and to read.

Responses following runs and alternation tuples.—The proportion of repetition responses following runs and alternations tuples of various lengths were tabulated from the data of Trials 101–300 for each group. The results are given in Fig. 3. Tuples of length greater than eight are omitted, as well as all data points based on fewer than 100 instances in the denominator. Figure 3a plots the data for runs of various lengths. The successive points on each curve give the values of $R'(0)$, $R'(00)$, $R'(000)$, $R'(0000)$, etc. The inverted bowl shape of the five middle curves is the well known Jarvik (1951) negative recency effect, or gambler's fallacy. In Fig. 3b are shown the data for the alternation tuples. The successive points on each curve give the values of $R'(0)$, $R'(10)$, $R'(010)$, $R'(1010)$, etc. As would be expected, the probability of a repetition response decreases as the length of the alternation tuple increases. Somewhat more striking is the fact that in all groups (except Group 90 for which the relevant datum does not appear) the decrement following the third light in an alternation tuple is greater than the decrement following the second light in such a tuple. This result is not an artifact of the tabulational procedure. Indeed, if each point included data only from those alternation tuples having *precisely* the length specified on the abscissa, an even greater difference in the decrements would have been obtained. Finally, by comparing the two panels of Fig. 3, it can be seen than an alternation tuple has a greater overt effect than does a run.

The variation over trial blocks in the response following runs and alternation tuples is exhibited in Tables 2 and 3. These tables list the differences between the R' values for tuples of lengths 2, 3, and 4, and the R' values for tuples of the next shorter length. The values of $R'(0)$ are included in Table 2 in order to permit calculation of the separate R' values entering into the several differences.

A negative difference in Table 2 indicates that the probability of a repetition response decreases as the length of a run increases, i.e., a negative recency effect. The lowest order differences, $R'(00) - R'(0)$, are negative over the first 50 trials, but thereafter are nearly all positive. The two higher order differences also increase over trial blocks. These results suggest that the negative recency effect adapts out over trials. This adaptation seems to be more rapid for conditions with stronger repetition tendencies in the light sequences.

A positive difference in Table 3 indicates that repetition response probability decreases as the length of the alternation tuple increases. Most of the values of $R'(0) - R'(10)$ are negative in the first block of 50 trials. Thereafter they assume fairly stable positive values except perhaps in the transfer trials for Groups 10–30. The values of $R'(10) - R'(010)$ are fairly constant over trial blocks. These latter values are also uniformly larger than the values of $R'(0) - R'(10)$,

TABLE 2

DIFFERENCES OF R' VALUES FOR RUNS OF VARIOUS LENGTHS

Trial Block	R' and R' Diff.	Group								
		10	20	30	40	50	60	70	80	90
1–50	$R'(0)$	13	27	45	45	55	62	72	79	89
	$R'(00)-R'(0)$	01	−04	−03	−09	−04	−08	02	−02	−01
	$R'(000)-R'(00)$			−08	−12	−08	−08	−02	00	01
51–100	$R'(0)$	06	25	41	48	59	70	78	89	91
	$R'(00)-R'(0)$	01	04	01	01	02	04	06	02	02
	$R'(000)-R'(00)$			−06	−13	−07	−02	00	00	00
	$R'(0000)-R'(000)$			(−05)[a]		(−03)[a]		(−02)[a]		
101–200	$R'(0)$	05	17	38	51	61	70	80	90	93
	$R'(00) - R'(0)$	02	00	07	03	06	04	04	01	01
	$R'(000)-R'(00)$			−09	−06	00	00	−01	00	00
	$R'(0000)-R'(000)$			−02		−03		−02		
201–300	$R'(0)$	04	16	35	49	62	75	84	90	94
	$R'(00)-R'(0)$	05	01	09	00	02	05	03	00	01
	$R'(000)-R'(00)$			−07	02	−02	03	00	01	01
	$R'(0000)-R'(000)$			−03		−07		00		
301–400	$R'(0)$	51	39	45	49	62	69	72	74	70
	$R'(00)-R'(0)$	−04	04	08	02	05	06	07	05	06
	$R'(000)-R'(00)$	03	−07	−01	01	−02	04	00	01	03
	$R'(0000)-R'(000)$			−02		−02		−02		
401–500	$R'(0)$	49	46	50	50	60	68	68	71	69
	$R'(00)-R'(0)$	−01	05	07	01	02	05	06	01	07
	$R'(000)-R'(00)$	00	−03	−02	−01	00	05	03	04	03
	$R'(0000)-R'(000)$			−03		01		−01		

Note.—Decimal points omitted.
[a] Calculated from the data of Trials 1–100.

confirming and extending the results of Fig. 3b.

Response following other stimulus tuples.—It is also of interest to examine the sequential dependencies for stimulus tuples other than runs and alternations. The statistic considered is the difference of R' values of tuples which are alike except for the leftmost (trialwise most remote) stimulus event. These differences thus measure the influence of the trialwise most remote event of the tuple on the current response.

The R' values have unequal reliability since the number of instances of a given tuple varies systematically with acquisition treatment. The R' values presented for 3-tuples are based on at least 150 instances in a block of 100 trials. However, six of the 4-tuples of Table 5 had pooled frequencies between 50 and 100 instances. In the transfer trials, of course, all tuples of a given length are equiprobable, and the R' values of tuples of lengths 2, 3, and 4 are based on approximately 1000, 500, and 250 instances, respectively, for a group of 20 *S*s. It should be noted that pooling data across the sequences within each condition introduces a bias to the extent that the number of instances of each tuple varies across sequences. However, the bias would appear to be negligible in the present results. Of the 24 R' values which were also computed from the individual *S*'s data (see below), the average magnitude discrepancy was .005.

The 2-tuple difference, $R'(00)-R'(10)$, may be obtained by adding corresponding values of $R'(00) - R'(0)$ from Table 2 and $R'(0) - R'(10)$ from Table 3.

TABLE 3

DIFFERENCES OF R' VALUES FOR ALTERNATION TUPLES OF VARIOUS LENGTHS

Trial Block	R′ Differences	Group								
		10	20	30	40	50	60	70	80	90
1–50	$R'(0)-R'(10)$	00	−01	−02	−06	−02	−12	05	−03	−01
	$R'(10)-R'(010)$	01	02	06	09	04	25	20		
51–100	$R'(0)-R'(10)$	00	01	00	01	02	06	15	07	10
	$R'(10)-R'(010)$	03	01	08	05	10	14	17		
	$R'(010)-R'(1010)$			(04)[a]		(05)[a]		(11)[a]		
101–200	$R'(0)-R'(10)$	00	00	03	02	06	08	10	04	08
	$R'(10)-R'(010)$	02	02	05	09	07	08	11		
	$R'(010)-R'(1010)$			04		06		05		
201–300	$R'(0)-R'(10)$	01	01	04	00	02	07	06	00	09
	$R'(10)-R'(010)$	01	00	05	07	07	07	18		
	$R'(010)-R'(1010)$			05		06		13		
301–400	$R'(0)-R'(10)$	−04	04	08	02	04	06	08	04	06
	$R'(10)-R'(010)$	01	10	10	10	10	09	13	13	17
	$R'(010)-R'(1010)$			05		07		10		
401–500	$R'(0)-R'(10)$	−02	05	06	00	01	06	06	01	08
	$R'(10)-R'(010)$	01	10	11	12	06	10	11	19	13
	$R'(010)-R'(1010)$			06		02		11		

Note.—Decimal points omitted.

[a] Calculated from data of Trials 1–100.

TABLE 4

DIFFERENCES OF R' VALUES OF PAIRED 3-TUPLES

Trial Block	Group								
	10	20	30	40	50	60	70	80	90
	(a) $R'(000)-R'(100)$								
1–50			−11	−19	−17	−17	−08	−02	06
51–100			−08	−21	−16	−05	00	01	03
101–200			−17	−09	−01	−02	−04	01	00
201–300			−09	03	−04	07	00	04	04
301–400	07	−13	−02	01	−03	07	00	03	06
401–500	01	−05	−03	−01	01	09	06	07	06
	(b) $R'(110)-R'(010)$								
1–50	12	08	18	20	09	35	30		
51–100	22	03	26	13	22	25	25		
101–200	15	06	19	19	13	13	16		
201–300	08	03	15	15	13	11	23		
301–400	01	18	18	19	18	17	24	24	30
401–500	01	18	21	21	11	19	20	18	24

Note.—Decimal points omitted.

These differences rise from predominantly negative initial values to level off at reasonably stable and definitely positive values of the order of .1. Except for Groups 10 and 20, there seems to be neither any systematic between-groups variation, nor any change in the transfer trials.

The two 3-tuple differences are presented in Table 4 for each group and for successive blocks of trials. The separate R' values in each difference may be readily evaluated by making use of the data of Tables 2 and 3. The values of $R'(000) - R'(100)$ are mostly negative in the early trials, but tend to become positive in the later trials for the groups with higher acquisition values of conditional probability. This trend presumably reflects the adaptation of the negative recency effect previously mentioned. The values of $R'(110) - R'(010)$, although suggestive of some differences between groups and over trial blocks, are rather the more remarkable for the degree of constancy which they exhibit. The average value is of the order of .2 which, being larger than the value of $R'(00) - R'(10)$, re-emphasizes the influence of alternation tuples on *S*'s response.

In order to obtain some indication of the variability of the scores discussed in the preceding two paragraphs, R' values were computed for the individual *S*s of Groups 30, 50, and 70, for Trial Blocks 101–200, and 401–500. The resulting analyses are given in Table 5. In particular, the *F*s for Mean show that the R' values differences, averaged over the two blocks of trials and over groups, are significantly different from zero.

The four 4-tuple comparisons were made only for Groups 30, 50, and 70; the results are given in Table 6. The values of $R'(0000) - R'(1000)$ are largely negative, again reflecting the negative recency effect. The values of $R'(0010) - R'(1010)$ are all positive and fairly large. Thus an alternation rather than a repetition three trials back in the sequence of lights still exerts a decided effect on the current response. The two remaining differences are essentially zero for Groups 30 and 50, but tend to be positive for Group 70.

The results of this section give ample evidence of the influence of the more remote stimulus events on the current response. In interpreting these results, however, it should be specifically noted that a nonzero difference implies nothing about *S*'s perception of, or memory for these more remote stimuli. Any model which takes account of the trial-to-trial changes in the response probability would predict nonzero differences (Anderson, 1959; Estes & Straughan, 1954).

Response–response dependencies.—Further information may be obtained by considering the dependence of the current response on the preceding response. Unfortunately, these dependencies are more difficult to interpret than are the stimulus dependencies because of a selection effect, and because of certain

TABLE 5

ANALYSIS OF VARIANCE OF R' VALUE DIFFERENCES FOR GROUPS 30, 50, AND 70

Source	*df*	*F* Ratios For		
		$R'(00) - R'(10)$	$R'(000) - R'(100)$	$R'(110) - R'(010)$
Mean	1	51.9*	4.81*	80.1*
Groups	2	1.18	5.42*	1.80
Error (b)	101	(.045)[a]	(.041)[a]	(.069)[a]
Trials	1	2.75	8.56*	0.23
T × G	2	3.40*	2.59	0.66
Error (w)	101	(.020)[a]	(.036)[a]	(.032)[a]

* $P < .05$.

[a] Error mean squares.

TABLE 6

DIFFERENCES OF R' VALUES OF PAIRED 4-TUPLES

Trial Block	Group 30	Group 50	Group 70	Group 30	Group 50	Group 70
	(*a*) $R'(0000)-R'(1000)$			(*b*) $R'(1110)-R'(0110)$		
1–100	08	−10	−07	05	00	04
101–200	−03	−06	−07	−02	01	12
201–300	−06	−12	00	00	03	05
301–400	−04	−04	04	−02	−05	09
401–500	−07	01	−02	01	00	06
	(*c*) $R'(0010)-R'(1010)$			(*d*) $R'(1100)-R'(0100)$		
1–100	10	11	17	07	−03	07
101–200	12	11	06	00	01	04
201–300	16	10	17	01	02	02
301–400	11	15	20	00	−01	03
401–500	11	03	20	−06	−03	−03

Note.—Decimal points omitted.

biases. These two difficulties will be discussed in the appropriate places.

The simple first-order response–response dependency is found by counting the frequency with which the key press on one trial is repeated on the next trial. This statistic was computed for each *S* of Groups 30, 50, and 70, over Trials 401–500, with resulting mean proportions of .538, .533, and .538, respectively. The three values do not differ significantly among themselves, but their mean is greater than the chance value of .500, *F* (*df* = 1, 101) = 17.6. This same statistic was also calculated for each of the ten 50-trial blocks for Group 50, but there was very little trial variation. The mean over all 500 trials was .538.

The result of the previous paragraph does not necessarily imply either that *S*s had a set or tendency to repeat the previous key press, or that the previous response has any determining influence on the present response; it could equally well represent the workings of the laws of probability according to the following argument. The probability of a left key press, averaged over trials on which a left key press occurred, will be greater than this same quantity averaged over trials on which a right key press occurred. It may then be expected (at least in the present case; see also below) that the probability of a left key press, averaged over trials following a left key press, will be greater than this same quantity averaged over trials following a right key press. The response will thus tend to repeat itself so that there is no present reason to believe that the obtained value of the response–response dependency represents anything more than this response selection effect. Any mathematical model would, of course, predict the magnitude of this response selection effect and thereby make use of the obtained data as a test of goodness of the model (e.g., Anderson, 1959; Anderson & Grant, 1958). It should be noted that, in general, some response selection effect will always be obtained whenever successive responses are not, for whatever reason, statistically independent. Consequently, the joint dependencies to be considered next are subject to the same difficulties of interpretation.

The dependence of the current response on the joint occurrence of the previous response and the four previous stimuli was also investigated. The notation for these dependencies is similar to that used for the stimulus dependencies. In the argument of each R' value, the symbols preceding the semicolon denote the previous stimuli in the same way as before, and the symbol following the semicolon denotes the previous response. Thus, $R'(1010;\ 1)$ denotes the proportion of repetition responses given that the last four stimulus lights were 1,0,1,0, in that order, and that the last response was 1. In each case, the last listed stimulus is the reinforcing stimulus for for the listed response. As before, the data were pooled over key positions so that, for instance, $R'(1010;\ 1)$ and $R'(0101;\ 0)$ denote the same quantity. There is thus no restriction involved in presenting the data in such a way that the last stimulus is always 0, and this convention has been adopted. Consequently, the R' values always denote the conditional probability of the response 0 on the next trial.

These R' values were obtained by pooling the data over *S*s and over trials for Trial Blocks 1–300 and 301–500. The comparisons presented in Table 7 are differences between R' values based on the same stimulus 4-tuple, and differing only in the previous response. Nonzero differences correspond, therefore, to predictability conferred by knowledge of the previous response over and above predictability gained by knowledge of the four previous stimuli. Table 7 is so arranged that the last response was correct in the R' value on the left side, incorrect in the R' value on the right side of each difference. In addition, the first four lines are for those cases in which the last two stimuli were alike, and the last four lines are for those cases in which the last two stimuli were unlike.

The problem of bias must be considered before discussing the results. Since the R' values are proportions conditional on the previous response, they involve a denominator which is subject to variation. This random variation introduces a bias which is approximately inversely proportional to sample size (Cramer, 1946, Sec. 27.7). An R' value computed for an individual *S* will thus be biased because of the limited number of responses made by that *S*. If *S*s were identical in the sense of having the same learning rates, this bias could (and should) be reduced by pooling the data of the group in order to compute a single ratio. However, if *S*s are not identical, this pooling procedure will itself introduce a bias because of the limited number of *S*s. This bias will be found even when there is sufficient data for each *S* to render negligible the bias in the individual ratios. In fact, the bias in the individual scores reported below is negligible; however, the entries of Table 7, which are based on pooled group data, do involve appreciable bias.

TABLE 7

DIFFERENCES OF PAIRED R' VALUES FOR STIMULUS-RESPONSE DEPENDENCIES

R' Differences	Group 30		Group 50		Group 70	
	Trials		Trials		Trials	
	1–300	301–500	1–300	301–500	1–300	301–500
$R'(0000;0)-R'(0000;1)$	51	37	37	34	34	35
$R'(1000;0)-R'(1000;1)$	21	26	20	15	22	22
$R'(0100;0)-R'(0100;1)$	04	09	13	21	15	12
$R'(1100;0)-R'(1100;1)$	07	10	10	21	11	04
$R'(1110;0)-R'(1110;1)$	−02	−02	−03	02	04	00
$R'(0110;0)-R'(0110;1)$	−05	04	−07	−07	−06	−01
$R'(0010;0)-R'(0010;1)$	−04	00	−06	−06	05	18
$R'(1010;0)-R'(1010;1)$	−07	01	−09	−04	−02	−11

Note.—Decimal points omitted.

This observed bias may, therefore, be taken as reflecting the existence of individual differences.

It may be noted, parenthetically, that any model which took into account the trial-to-trial changes in response probability would yield theoretical expressions (possibly of a Monte Carlo variety) for the *unconditional* proportions. The use of such data would avoid the problem of bias arising from the use of ratios with a random variable denominator. Since the two models considered here have already been seen to be inadequate, this procedure has not been followed. Without a model, however, the unconditional proportions are uninterpretable. Consequently, the nonparametric procedure adopted here, despite its defects, seems necesary to obtain meaning from the results.

The results for Group 50 will be considered first since they are the most reliable. The entries in the first four lines of Table 7 are positive and large. Positivity would be expected on the basis of the response selection effects discussed above since the previous key press in the R' values on the left of each difference is a 0, and since the repetition response is a 0 for both R' values of each difference. However, the Table 7 entries seem considerably larger than one would expect if the response selection effect alone were operating. The most appropriate measure of the size of the response selection effect is obtained by doubling the difference between the simple first-order response–response dependency and the chance level. The previous data of this section thus yield the value of .076 against which to compare the Table 7 entries. This comparison value may actually be too large since taking into account the four previous stimuli, as is being done here, would be expected to reduce the amount of information conferred by knowledge of the previous response.

In order to test this comparison, the bias in the entries of Table 7 must be allowed for. Two values of $R'(000;\ 0) - R'(000;\ 1)$ were calculated for each S of Group 50, each value based on the data of alternate 50 trial blocks. Since the bias is inversely proportional to the amount of data, comparing the mean of these two values with the corresponding value based on all the data for that S yields a measure of the bias in the individual scores arising from the use of the ratio. This bias was negligible. It may be noted, incidentally, that the correlation between the two scores was .55. The mean of the individual values of $R'(000;\ 0) - R'(000;\ 1)$ was .24 which is significantly larger than the comparison figure of .076, $F\ (df = 1, 39) = 50.83$. Also, the .24 value is less than the weighted mean of .27 obtained from the Group 50 entries in the first two lines of Table 7. It is thus seen that these particular entries are biased away from zero.

The finding that the response selection effect is insufficient to account for the Group 50 entries in the upper half of Table 7 should be interpreted cautiously. It may be necessary to allow for a determining effect of the previous response on the present response as is done in the experimenter–subject controlled models of Bush and Mosteller (1955). A second possibility is that the assumption of path independence must be abandoned as in the models proposed by Sternberg (1959b).

The Group 50 entries in the lower half of Table 7 are predominantly negative. Statistical significance was investigated by computing individual values of $R'(10;0) - R'(10;1)$ according to the procedure used for the upper half of the table. The split-half analysis yielded a correlation of .67, and showed also that the bias in the individual scores was negligible. The mean of the individual scores was −.036, with a standard error of .030. The tabular entries are again biased away from zero, since the corresponding mean obtained from Table 7 was −.053. Although the Group 50 entries in the lower half of Table 7 are not significantly less than zero, they deserve comment since the response selection effect noted above would, if taken at face value, imply that these entries should be somewhat greater than zero. In accord with the original model, and with the preceding analysis of stimulus dependencies, it is assumed that *S*s have an alternation response (predicting next that light which did not flash last) in their repertoire. Now, in the R' value on the left of each difference, an alternation response (a 0) has just been made (and rewarded). But in the R' value on the right side of each difference, a repetition response has just been made (and punished). Hence it would be expected on purely probabilistic grounds that an alternation response (a 1) would be more likely on the next trial in the first case than in the second. This means that a repetition response on the next trial will be less likely in the first case than in the second, in agreement with the data. By thus interpreting the overt key presses as alternation and repetition responses, it is seen that the simplest explanation of the negative values is in terms of a response selection effect. Conversely, the fact that the entries are not positive gives added support to the existence of the alternation response. The further interpretation of the results as indicative of differential effects of reward and nonreward would be premature.

The data for Groups 30 and 70 need little discussion. The entries in the upper half of Table 7 present essentially the same picture as Group 50 did. The data for the lower half of the table do not show such good agreement. This might perhaps be expected for Group 70 since the alternation response presumably does not develop as strongly there. For Group 30, individual values of $R'(10;0) - R'(10;1)$ over Trials 1–300 were computed using the split-half technique as before. The correlation of the two scores was .69, and again the bias in the individual scores was negligible. The mean individual score was $-.061$ with a standard error of .031. While the mean thus falls slightly short of being significantly less than zero, it reinforces the corresponding analysis of the Group 50 data. It should be noted, however, that the values over Trials 301–500 for Group 30 are not negative, a result which is contrary to the above explanation.

Miscellaneous Results

This section gives two results from the combined data of Exp. I and II.

It has been observed that Groups 30 in both experiments show a hump at the third block of 10 acquisition trials, first increasing and then decreasing in frequency of repetition responses. Tests were made on the difference of the total scores on Trials 1–10, and Trials 21–30. The mean difference was significantly greater than zero both for Exp. I, $F\ (df = 1,\ 48) = 18.81$, and for Exp. II, $F\ (df = 1,\ 24) = 12.46$. Moreover, there were no complications from other factors: in Exp. I, neither trial rate, stimulus sequences, nor their interaction were significant; in Exp. II, where eight different sequences were used, the F for Sequences was again close to unity. Group 40 lay significantly below the 50% level on the first block of 10 trials, which is consistent with the behavior of Group 30.

These results suggest the presence of an initial tendency toward alternation responses. In order to obtain a purer measure of initial tendencies, the data from the first two key presses are given in Table 8. Here the first stimulus light follows the first key press and precedes the second key press. Table 8 includes the present data, that from Exp. 3 of

TABLE 8

JOINT DISTRIBUTION OF FIRST TWO KEY PRESSES AND FIRST REINFORCING STIMULUS

First Two Key Presses	First Reinforcing Stimulus		Total
	Right Light	Left Light	
Right-Right	119	114	233
Right-Left	99	56	155
Left-Right	40	44	84
Left-Left	85	47	132
Total	343	261	604

Note.—The reinforcing stimulus follows the first key press and precedes the second key press.

Anderson and Grant (1957), and that from a pilot study, to yield a total *N* of 604. In all cases, *S*s had three previous practice trials in which both lights or neither light could come on, but not either light singly. There is a significant preference for the right key on Trial 1, but not on Trial 2. The probability that the second press is an alternation response is .57, significantly above chance by a binomial test. It will be noted that the data could also be interpreted in terms of the "win–?, lose–stay" strategy of Goodnow and Pettigrew (1955). A somewhat different treatment of these data has been made by Sternberg (1959a) in terms of a model embodying a perseverative tendency.

DISCUSSION

This discussion will first compare the present results with those of other experiments, and then consider the mathematical model approach to probabilistic learning situations of the kind investigated here.

The most relevant comparison data are those of Engler (1958). Four conditions in her Exp. II had acquisition values of π_{11} identical with those of the present Groups 10, 50, 70, and 90. The numerical values of the terminal acquisition levels are in fairly good agreement except that, for Group 10, the present asymptote of .04 is considerably lower than the .11 figure obtained by Engler. Her Group G, with $\pi_{11} = .25$, also terminates somewhat higher than might be expected by interpolation in Table 1. The source of this discrepancy is not clear. However, it is somewhat disturbing in view of the generally good agreement among different experiments on asymptotic values for the frequency variable.

It is worth mentioning that Groups 0 and 100, which received continuous reinforcement in acquisition, were little different from Group 50 in the transfer trials, but had significantly different transfer asymptotes than the other alternation and repetition groups. This result is not inconsistent with the common finding of greater resistance to extinction after partial reinforcement if the transfer condition is considered as analogous to extinction. It is possible, of course, that Groups 0 and 100 would have reached different transfer asymptotes had they been given more than 100 acquisition trials.

The finding that Group 0 learned faster than Group 100 is the reverse of the result of Goodnow and Pettigrew (1956). A plausible explanation lies in the difference in the reinforcing stimulus situation. In their "two-armed bandit" task, the fact that the other response would have been correct on trials when *S* received no chip was established by instruction, but no special signal to this effect was given during the experiment as was done in the procedure here.

The results from the alternation tuples do not seem to be entirely in agreement with corresponding results of Nicks (1959). In particular, $R'(0010) - R'(1010)$ has the value of about .2 for Groups 30, 50, and 70, (Table 6). Nicks' Table 2, however, shows values of the same statistic of .02 and .04. The tabulation procedures were the same in both cases, and the reliabilities of the data are such as to make the discrepancy significant. Nicks reports these data only for his 67:33 frequency condition so that the source of the discrepancy may stem from this difference in the experimental situations. In any event, the present data show that Nicks' emphasis on runs as the only important stimulus tuples is not justified in general.

The final comparison with existing data is restricted to Group 50. For this condition, a statistic obtained by doubling the difference between the repetition response level and the chance level is the same as a sequential statistic computed by Estes and Straughan (1954). This quantity has here an asymptotic value of about .24 which agrees with the average value of .24 obtained from their three frequency conditions. In the report of Anderson and Grant (1957), however, these values (there denoted by *a*) tended to be somewhat larger. In all three of these experiments it was found that the *a* values were considerably lower in the early trials.

The predictions made by the present model, as well as by the revised model of Burke and Estes (1957), disagree with the experimental results in a number of respects. In the first place, the acquisition asymptotes at the lower π_{11} values are considerably below the theoretical values. Second, the analyses of the sequential dependencies have exhibited several deviations: the negative recency effect, which is commonly reported, was again found here; the alternation tuple data showed that the third preceding light in the stimulus tuple, *RLR*, had a greater effect than the second preceding light in the stimulus tuple, *RL*, a result which is contrary to both models; in addition, the analysis of the response dependencies suggested that knowledge of the previous response yielded more information about the present response than would be expected theoretically, at least for models similar to those considered here. Finally, the fact that the final levels in the transfer condition (in which all groups received the same treatment) were different, and showed no sign of converging, is definitely contrary to the models.

Although these discrepancies have been obtained in the conditional probability situation, it is evident that they are in large part applicable to models for the frequency situation as well. In particular, the $\pi_{11} = .5$ condition is common to both the frequency and the conditional probability situations. Thus, although the conditioning process assumed in current mathematical models for the two-choice task may well occur, it would appear that the models neglect much of the underlying processes. Furthermore, since the mean learning curve comprises an average of the sequential increments and decrements in response probability, the agreement between obtained and predicted "matching solution" behavior must be considered fortuitous.

A necessary condition for constructing an adequate model is brought out by the sequential dependency analyses. The negative recency effect and the alternation tuple results imply specifically that it will be necessary to take into account the reinforcing stimulus on more than the one preceding trial which was considered in the two models discussed here. The situation could then be considered as discrimination learning, with memory traces of the last several events serving as discriminative stimuli in a system of S-R associations. It would probably be desirable to allow an analogous trace representation of preceding responses in the model, although the existing evidence is not unambiguous on the necessity for this.

The formulation of the preceding paragraph is consistent with the claim of Hake and Hyman (1953) that *S*s perceive and respond to specific sequences of events preceding each trial, and is perhaps not opposed to the notion of "strategies" used by Goodnow and Pettigrew (1955). However, it is apparent that the elucidation of the effective discriminative stimuli will not be as simple as has been thought. The specification of these stimuli will probably rely heavily on sequential dependency data, but these, as has been seen, include a number of automatic effects not directly indicative of underlying psychological processes. Indeed, it would appear that the dissection of these data cannot be carried very far without an exact model. Estes' (1950) conceptualization of the stimulus, together with the treatment of probabilistic discrimination learning given by Burke and Estes (1957), affords a possible method of attack. Thus, for instance, shorter tuples or tuples of greater perceptual impact would correspond to larger subsets of stimulus elements or to subsets of stimulus elements with greater sampling probabilities. The stimulus element approach would also be well suited to quantifying the similarity of the total stimulus trace on successive trials. Sternberg's (1959b) path-dependent models with many-trial dependencies represent a second possible approach to the problem.

It was expected that the sequential

dependency analyses would give the most incisive information about the behavior since they most directly reflect the trial-to-trial increments and decrements in the variables underlying the overt response. Nevertheless, the portion of these data which was immediately useful was not overly large. As has been seen, the nonparametric interpretation of such data is difficult when the behavior is as complicated as is the case here. The sequential dependencies are most useful in testing or evaluating parameters of particular models since any quantitative model must predict their magnitude. A general mathematical treatment of sequential dependencies for linear learning models has been given by Anderson (1959). It can be shown, for example, that the original Estes and Straughan (1954) model implies that the entries within each of Tables 5 and 6 should be constant and equal to $\theta(1-\theta)^2$, and $\theta(1-\theta)^3$, respectively. Corresponding expressions for the two models discussed here were not worked out since they failed on other grounds. However, the cited expressions point up the usefulness of the dependency data. It is hoped that the present results will be of use in testing future models.

The behavior in the transfer condition raises a serious problem for any theoretical treatment of the two-choice probabilistic situation. It will be recalled that in the transfer condition, in which all *S*s received a common treatment, the various groups showed no sign of converging to a common level of repetition responses even after hundreds of trials. The data also indicated that this transfer effect was not smoothly related to acquisition treatment. From a qualitative point of view, it is apparent that the acquisition treatment has induced some permanent or semipermanent change in the *S*s. However, the results of the sequential dependency analyses have shown that at least some of the various stimulus tuples must be considered separately. Consequently, since the induced change could be associated with any or all of these tuples, it is not possible to get a clear specification of its locus on the basis of the existing data.

The same problem holds for the mathematical model approach, but in an even more acute way since quantitative results are desired. The trace stimulus formulation would be feasible if the traces are short range and do not extend back more than, say, a half dozen trials. However, the transfer behavior is a long-range effect spanning several hundred trials. It is clearly impracticable to allow directly for such long term traces in any model. The same difficulty would apply to Sternberg's (1959b) path-dependent models.

It is possible, of course, that the long range effects arise from systematic changes (e.g., Restle, 1955) in the populations of stimulus elements which give rise to the short range traces. If this is correct, the suggested development would still be practicable although it would require the incorporation of the laws of such stimulus change. There is some evidence which may very gingerly be interpreted as suggesting that such laws of stimulus change may not be too complicated. The fact that the between-groups differences in the higher order dependencies were not more marked indicates that the changes in the stimulus element populations may possibly be largely confined to the stimulus 2-tuples. However, such a simple resolution of the problem of long range effects would still be insufficient to salvage the two models discussed here.

It seems not unreasonable to suggest that the most promising present line of attack on the two-choice situation lies in an extensive empirical investigation of the behavior under a variety of transfer conditions. This would include the study of repetition responses in situations in which the light frequencies are the independent variables. The present data, and those of Engler (1958) and of Goodnow and Pettigrew (1955) who also used a transfer paradigm, cover only a small part of the picture. Until more knowledge of transfer behavior is available, spot checking of specific models is liable to be little better than shooting in the dark.

In future work, it might well be more efficient to procure large quantities of data for a few *S*s so that reliable individual values of the sequential dependencies can be obtained. It is apparent from the present results that a number of parameters will be needed to account for the behavior, whether or not this is done in a formal mathematical way. Averaging data over *S*s in order to increase reliability is only too likely to simultaneously confound the descriptive statistics with individual differences (Anderson, 1959). This increases the difficulty of interpreting the statistics as well as any model parameters derived from them.

SUMMARY

Two experiments on two-choice probability learning were reported which varied the first-order conditional probability in the stimulus event sequences. This variable is the probability that an event occurs on one trial, given that it occurred the previous trial. The dependent variable was predictions of that same event which had occurred the previous trial.

Experiment I used a 3 × 2 design with three trial rates and two values of conditional probability in the 200 acquisition trials. Acquisition was followed by transfer to completely random sequences. The conditional probability groups reached significantly different acquisition asymptotes, and stable, but significantly different, response rates in transfer.

In Exp. II, 11 groups were given acquisition training with conditional probabilities ranging from 0 to 1. Excepting the two extreme groups, 300 acquisition trials, 200 transfer trials, and 60 recovery trials were given. Acquisition asymptote was an orderly function of conditional probability. Stable but unequal response rates in transfer were obtained, as in Exp. I. Spontaneous recovery was also obtained.

Detailed analyses of sequential dependencies in the stimulus-response sequences found, in addition to the "gambler's fallacy," strong effects of strings of consecutive alternations in the stimulus sequences. Some evidence for a determining effect of previous responses was obtained.

Two models based on statistical learning theory predicted asymptotic acquisition response levels quite well for the higher values of conditional probability, but did poorly at the very low values. The models also failed to account for the nonconvergence in transfer, and for much of the sequential dependency data. It was concluded that agreement between obtained and predicted "matching solution" behavior is fortuitous.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. Unpublished doctoral dissertation, Univer. of Wisconsin, 1956.

ANDERSON, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959.

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.

ANDERSON, N. H., & GRANT, D. A. Correction and reanalysis. *J. exp. Psychol.*, 1958, **56**, 453–454.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

CRAMER, H. *Mathematical methods of statistics.* Princeton: Princeton Univer. Press, 1946.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **2**, 1–42.

ENGLER, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GOODNOW, J. J., & PETTIGREW, T. F. Effect of prior pattern of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, **49**, 381–389.

GOODNOW, J. J., & PETTIGREW, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51**, 385–392.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

STERNBERG, S. H. An analysis of four models for human learning. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press. (a)

STERNBERG, S. H. A path-dependent linear model. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. (b)

WILK, M. B., & KEMPTHORNE, O. Fixed, mixed, and random models. *J. Amer. Statist. Ass.*, 1955, **50**, 1144–1167.

(Received May 4, 1959)

Journal of Experimental Psychology
Vol. 59, No. 2, 1960

RELATION OF SERIAL POSITION ERRORS TO DOUBLET AND SPLIT-DOUBLET LOCATION IN VERBAL MAZE PATTERN[1]

RONALD L. ERNST, DONALD R. HOFFELD, SIDNEY SEIDENSTEIN, AND W. J. BROGDEN

University of Wisconsin

Thompson, Voss, and Brogden (1957) investigated the effect of doublet location within the verbal maze pattern on its acquisition and the form of the serial position error curves. The 16-unit 4-choice pattern of Brogden and Schmidt (1954) was modified by locating doublets (two successive like responses) at Units 2 and 3, 5 and 6, 8 and 9, 11 and 12, and 14 and 15, respectively, for the five experimental groups. The serial position curves of the difference in errors between each experimental group and the control group show a doublet effect of a significant increase in errors for the second item. Analysis of first-order contingent responses on the first trial shows that *S*s tend not to repeat a response correct for the immediately preceding position and suggests that this kind of guessing behavior during the discovery phase of acquisition accounts for the doublet effect.

The method of producing a doublet for the experimental group patterns generated split-doublets (two identical response items with a different response item between them), of which there were none in the control group pattern. The split-doublets produced a significant increase in error at the position of the second like item (third position of split-doublet sequence). The unintended confounding of doublets with split-doublets in the experimental maze patterns may be responsible for the results. Of the two experiments to be reported[2], one is an exact repetition of the one performed by Thompson, Voss, and Brogden (1957) with elimination of split-doublet sequences from the patterns of the experimental groups. The other experiment was designed to study the effect of location of split-doublets.

PROCEDURE

Experiment I.—The 16-unit maze with the four alternate choices, 10, 20, 30, and 40, each occurring four times in the pattern, was learned by the control group. The five patterns acquired by the experimental groups contained a single doublet at Maze Units 2 and 3, 5 and 6, 8 and 9, 11 and 12, 14 and 15, respectively, and each represented a limited modification of the standard maze to provide a doublet without split-doublets or other noticeable sequences elsewhere in the maze. Each maze was permuted to obtain four subpatterns and thus provides for each number to be represented once at each of the 16 positions.

Of 24 *S*s in each group, six learned one of the four subpatterns. Half of the *S*s in each condition were treated by one of two *E*s. The *S*s were students in elementary psychology at the University of Wisconsin who volunteered to serve. Assignment of *S*s to conditions was random within each replication.

The instructions of Thompson, Voss, and Brogden (1957) were read to *S*. Responses of 10, 20, 30, or 40 were made by *S* with *E* signaling the correct response by saying "Forward." This procedure continued until *S* completed one errorless trial. Each response at each unit was recorded by *E* for each trial. A stop watch was used to measure the time of each trial and the 1-min. rest between trials. Other conditions were comparable to those of Thompson, Voss, and Brogden (1957).

---

[1] This research was supported in part by grants from the National Science Foundation and the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

[2] Hoffeld and Seidenstein collaborated with the senior author on Exp. I and Ernst on Exp. II.

Experiment II.—The four patterns acquired by the experimental groups contained a single split-doublet at Maze Units 2–4, 5–7, 10–12, and 13–15, respectively, and each represents a limited modification of the standard maze to provide a split-doublet without doublets or other noticeable sequences elsewhere in the maze. Except that a single *E* treated all *S*s (24 in each group, 6 in each subgroup), the procedure for Exp. II was identical to that for Exp. I.

RESULTS

Experiment I.—Analyses of variance of the data in terms of time and trials to the criterion trial showed no significant effect for maze pattern, maze pattern subgroups, *E*, or the interaction of these variables. Since the assumption of homogeneity of variance was found not tenable for the error data, the 1% level was used to determine statistical significance for all analyses of these data. No source from the between-*S*s variation was significant. For the within-*S*s variance, the significance of serial position establishes the reliability of the serial position error curves. Of the significant interactions, only that for serial position by maze pattern is meaningful in terms of the experimental design. Graphic plots of the other interactions (serial position by submaze pattern and serial position by maze pattern by submaze pattern) show differences that appear to be only of random character.

Since there are no significant differences between groups in terms of total errors to the criterion, and since the primary results are the variation in form of the serial position error curve as a function of maze pattern, the lead of McCrary and Hunter (1953) was followed by converting the error data at each serial position to percentage of total error. The number of errors over all trials to the criterion was converted to percentage of total error over all trials and all positions for each of the 16 serial positions for each *S*. The group mean of the percentage scores for *S*s of each experimental group was computed for each serial position and the plot of these values is presented in Fig. 1 by the solid line curves. In order to increase the stability of the standard serial position error curve, data for the control group of Thompson, Voss, and Brogden (1957) were converted to percentages and combined with the data for the present control group. Analysis of variance of the data for the two control groups shows no significant differences between them. Comparison of the serial position curve for all errors in the combined control group (Fig. 1) with the comparable serial position curves for the experimental groups shows substantial differences in the form of the curves. These differences are presented more clearly in the solid line curves in the left half of Fig. 2. The data for these curves were obtained by taking the difference in percentage of error for each *S* and the mean percentage error for the combined control group at each of the 16 serial positions and obtaining the mean of these differences for each of the five experimental groups. For each group there is an increase in error for the second position of the doublet. The difference data were subjected to an analysis of variance resulting in a significant interaction of serial position by maze pattern. This establishes significant differences in the form of the differ-

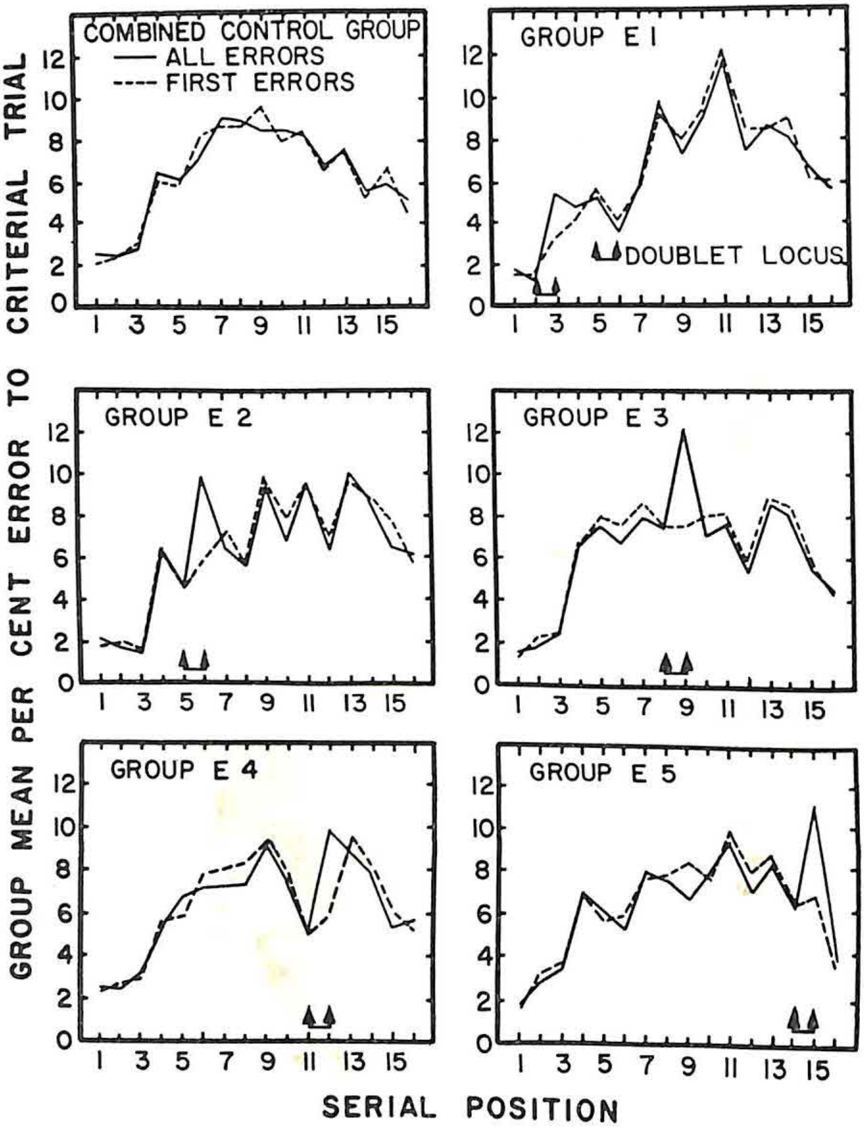


FIG. 1. Serial position curves in terms of percentage of total errors and percentage of first errors as a function of doublet locus.

ence curves between the five experimental groups. The differences in the form of the curve involve the effect of doublet location and other differences attributable to variation in maze pattern required for construction of doublets at appropriate positions for the different experimental groups. The results shown by these curves are comparable to the results of Thompson, Voss, and Brogden (1957) for doublet location and are not confounded by the split-doublet effect that occurred in the earlier study.

Thompson, Voss, and Brogden (1957) demonstrated that frequency of first-order contingent response was reduced for any given number if it was followed by the same number.

The magnitude of this effect was approximately the same for the numbers 20, 30, and 40, but was substantially less for the number 10 (more responses of 10 when the number previously correct was 10). Analysis of first-order contingent responses for *S*s of the current study produced similar results. Serial position error curves based on the first errors per

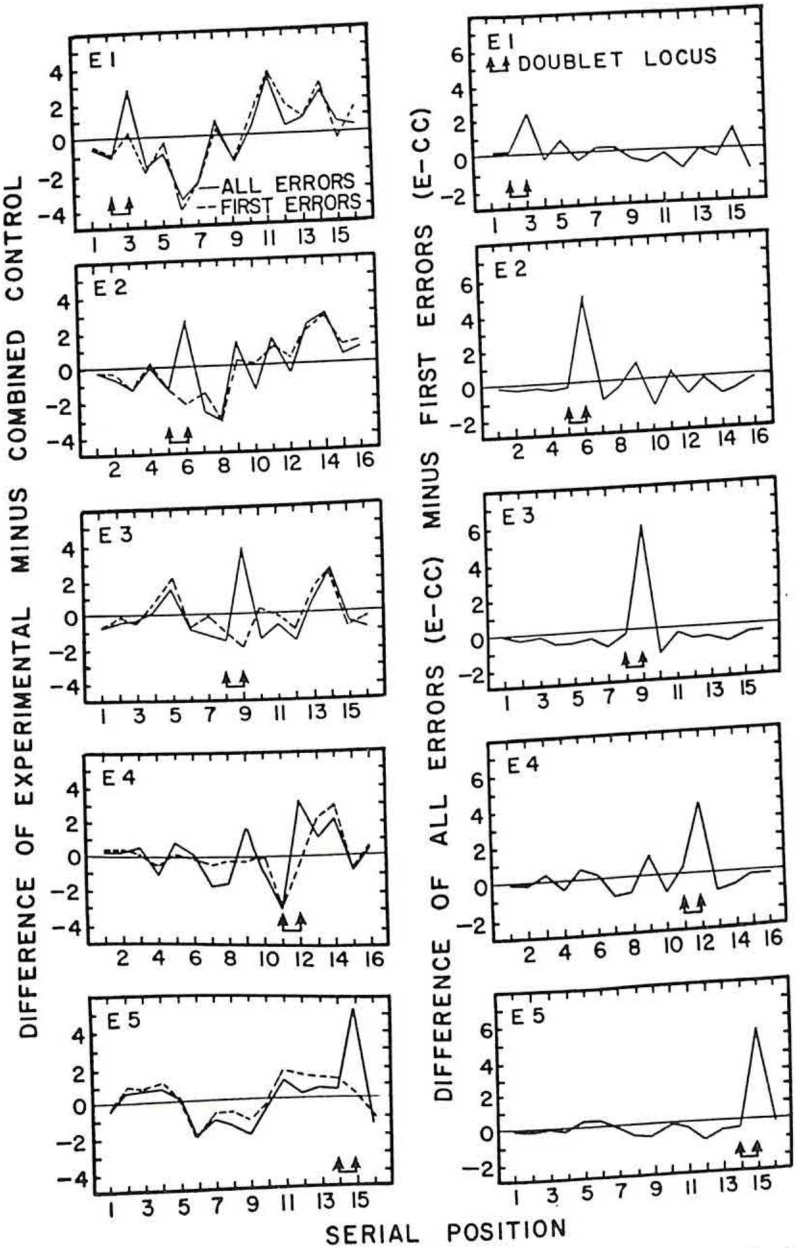


FIG. 2. Curves of differences between serial position curves for experimental minus control groups for percentage of total and first errors (left side) and curves of the differences between these difference curves for each experimental group (right side).

trial should show the doublet effect and provide more information about the relation between the doublet effect and the tendency not to repeat a response previously correct. The totals of the first errors for each position for each *S* were converted to percentages of the total first errors over all serial positions. The mean of the percentage values for each serial position was obtained for each of the six separate groups, and the broken line curves of Fig. 1 are based upon these data.

There is little difference in form of the serial position error curves based on all errors and first errors except for the five experimental groups at the position of the second item of the doublet. In the left half of Fig. 2 the broken line curves represent the difference between the first-error serial position curve of the experimental group and the first-error serial position error curve of the combined control group. The curves of the right half of Fig. 2 represent the difference between the two difference curves shown in the left half of Fig. 2. The only differences substantially greater than zero are those occurring at the second position of the doublet. Analyses of variance of the first-error difference data, and of the difference between the experimental and combined control group differences for all errors and for first errors show the interaction of serial position and maze pattern to be significant. The significant differences between the dotted line curves of the left half of Fig. 2 (which exhibit no doublet effect) indicate that the maze patterns represented by the experimental groups may produce a different form of the serial position error curve at other loci than those for the doublet. The significant differences between the curves of the right half of Fig. 2 apply only to the doublet effect and indicate that this effect is a direct function of doublet location for all-errors measure of the serial position error curves.

The *t* test was used to establish fiducial limits at the 1% level for mean differences between the percentage total error of experimental minus control (solid line curves, left half, Fig. 2), the percentage first error of experimental minus control (broken line curves, left half, Fig. 2), and the first of these minus the second of these differences (solid line curves, right half, Fig. 2). These limits are, respectively, ±2.8, ±2.8, and ±2.2. The second item of the doublet exceeds the positive limit for Groups E_3, E_4, and E_5 for the total-error measure, and for all groups for the difference between total and first errors. It is within .2 of equaling the positive limit for Groups E_1 and E_2 for the total-error measure. In the case of the first item of the doublet, that for Group E_4 exceeds the negative limit for both the total-error and first-error measures. There are other positions at which the various measures exceed the fiducial limits. These are as follows: for the total-error measure in Group E_1, Positions 6 and 7 (negative), and Position 11 (positive), and in Group E_2, Position 8 (negative) and Position 14 (positive); for the first-error measure in Group E_1, Positions 11 and 14 (positive), in Group E_2, Positions 8 (negative) and 14 (positive), and in both Groups E_3 and E_4, Position 14 (positive). These differences may be a complex function of the change in pattern relative to the pattern of the control group required to locate the doublet at the desired positions. The differences at Position 8 for Group E_2 are the only differences that occur at a position where the correct response is different from that of the control group. The only consistent relation is the increase in error at Position 14 for the first-error measure for all groups except E_5, in which group Position 14 is the first item of the doublet.

The failure to find any doublet effect with the first-error data is inconsistent with the demonstrated tendency of *S*s not to repeat a previously correct response and with the hypothesis that the doublet effect is primarily a function of this kind of guessing behavior during the discovery phase of acquisition of the maze. The doublet effect may be present in the first-error data during the discovery phase, but not show up in the data for all trials, either because of its small magnitude or because of a negative doublet effect after the discovery phase. To test this hypothesis the number of first errors for the first quarter and the last three-quarters of trials to the criterion was tabulated for each *S* for the doublet positions for the experimental groups. Similar data were tabulated for the control group at the serial positions for which doublets occurred in the experimental groups. The frequency of first error at the first doublet item was subtracted from that for the second item. From this difference was subtracted the mean difference for the appropriate serial positions for *S*s of the combined control group. This mean over all experimental groups is .7 for the first one-quarter of trials and −.5 for the last three-quarters of the trials. The results of one-tailed *t* tests show the mean difference for the first one-quarter of the trials to be significantly greater than zero and that for the last three-quarters of the trials to be significantly less than zero at the 1% level.

The total errors for the first quarter and last three-quarters of the trials were treated in the same way as the first-error data. The group mean difference over all experimental groups was 4.9 for the first one-quarter and 2.5 for the last three-quarters of trials. One-tailed *t* tests show both means to be significantly greater than zero at the 1% level, and the mean of 4.9 to be significantly greater than the mean of 2.5.

A further analysis of the doublet effect was attempted by constructing serial position curves from the trial data. The number of trials to and including the last error for each position was tabulated for each *S* and was converted to percentage of trials to the criterion. Examination of curves for group means and for mean differences between each experimental and control group shows them to be of similar form to the curves constructed from errors. There is no evidence of any doublet effect for Groups E_1 and E_5, but for Groups E_2, E_3, and E_4 there is a decrease in the number of trials for the second item of the doublet. No further analyses were made of the trial data.

Experiment II.—The trials, time, and error data of the four experimental groups and the combined control group (Exp. I) were subjected to analysis of variance. Since in each case the assumption of homogeneity of variance was not tenable, the 1% level was used to determine statistical significance for these and all subsequent analyses. There were no significant differences for the trial or time data. In the analysis of the error data, no source from the between-*S*s variation was significant. For the within-*S*s variance, the significance of serial position establishes the reliability of the serial position error curves. Of the significant interactions, only that for serial position by maze pattern is meaningful. Graphic plots of the other significant interactions (serial position by submaze pattern, and serial position by maze pattern by submaze pattern) show differences that appear to be of random character only.

Further treatment of these data follows that of the error data in Exp. I. Analyses of variance were completed on percentage of total error, difference in percentage of total error between the experimental groups and the control group, percentage of first

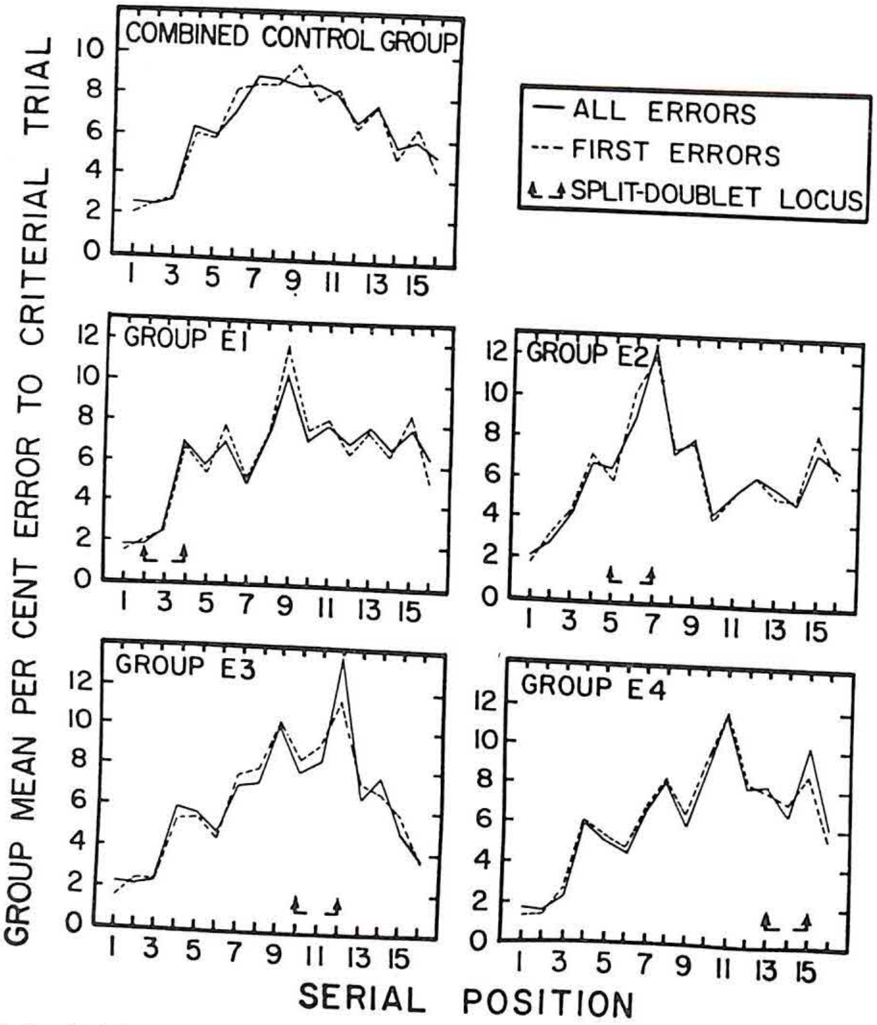


FIG. 3. Serial position curves in terms of percentage of total errors and percentage of first errors as a function of split-doublet locus.

error, and difference in percentage of first error between the experimental groups and the control group. The interaction of maze pattern and serial position was significant for each of these analyses. The differences produced in the form of the serial position error curve by the split-doublet are presented graphically in Fig. 3 and 4. The solid line curves represent total errors and the broken line curves represent first errors.

Examination of these curves shows no effect of the split-doublet for Positions 2–4 (E_1), an effect of increase in error at the final position of the split-doublet sequence for Groups E_2, E_3, and E_4, an increase in error for the second position of the split-doublet sequence for Groups E_2 and E_4, and no differences in curve form between the total error and first error measures. In addition, there are substantial differences between the experimental and control group serial position curves at positions other than the locus of the split-doublet.

The *t* test was used to establish fiducial limits at the 1% level for mean differences in percentage of total error between the experimental groups and the control group. The magnitude of positive difference in error (in favor of the experimental group) exceeds the fiducial limit for the third item of the split-doublet for Groups E_2, E_3, and E_4, and for the second item for Groups E_2 and E_4. The other positions at which the error differences exceed the fiducial limits are as follows: Group E_1, Position 7 (negative) and Positions 9, 14, and 15 (positive); Group E_2, Positions 10, 11, and 13 (negative), and Positions 3, 15, and 16 (positive); Group E_3, Positions 6, 13, and 16 (negative), and Positions 9 and 14 (positive); Groups E_4, Positions 6, 7, and 9 (negative), and Position 11 (positive). There are no consistent relations between these differences and specific positions or other features of maze pattern. They may be a complex function of the change in the maze pattern required to locate the split-doublet at the desired positions in the patterns of the experimental groups.

Further analysis of the split-doublet effect involved examination of the first-order contingent responses for the second and third split-doublet positions on the first trial. The results of χ^2 tests show a reduction in the frequency of response for a given number at the second position of the split-doublet sequence if it was the correct response for the first position. There are no differences in frequency of response at the third position of the split-doublet sequence as a function of the correct response for the first position. These results are comparable to those of Thompson, Voss, and Brogden (1957) and confirm their suggestion that the guessing habits of *S*s during the discovery phase do not account for the split-doublet effect. Examination of the first error data by quarters of trials to the criterion shows that the split-doublet effect is present in the first error data in each one of the four quarters of trials and declines in

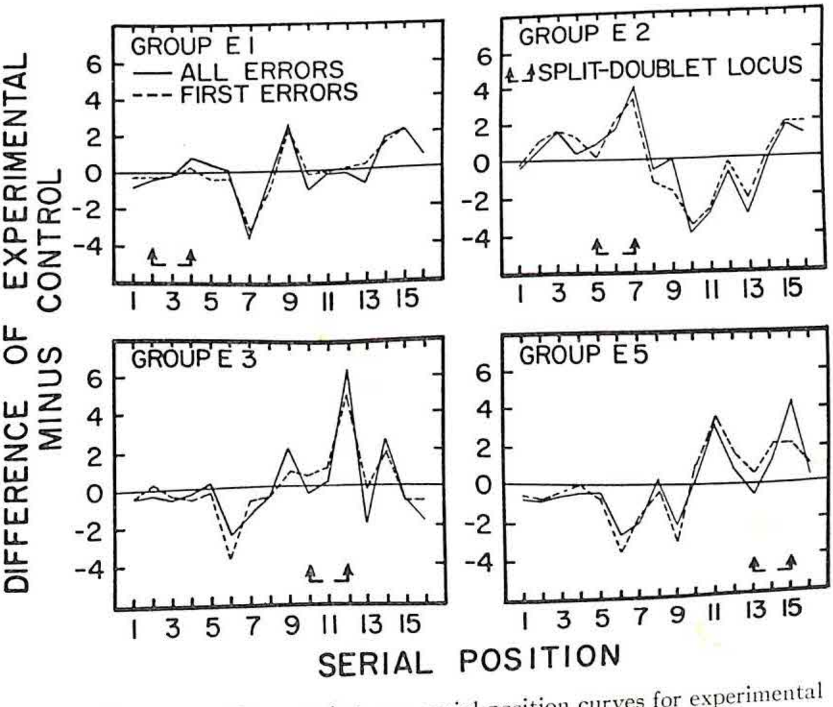


FIG. 4. Curves of differences between serial position curves for experimental minus control groups for percentage of total and first errors.

magnitude progressively from the first to the last quarter.

Discussion

The results of the two experiments establish a doublet effect and a split-doublet effect that are independent of each other. The doublet effect occurs at all locations in the maze pattern where tests for it were made, and there appears to be no difference in its magnitude as a function of the doublet locus. The split-doublet effect is differential in magnitude and in character as a function of its locus. The two phenomena are also different in that the doublet effect differs in the discovery and acquisition phases whereas the split-doublet effect does not. Both in the case of maze patterns involving doublets and split-doublets, there are other differences in form of the serial position error curves that may not be accounted for by the location of the doublet or split-doublet or by the interchanges necessary to their formation.

The doublet effect is evident for total errors over all trials, and for total errors for the first quarter and for the last three-quarters of the trials. Its magnitude is twice as great for the first quarter of trials as for the last three-quarters. The effect is not measurable by the first-error measure over all trials. For the last three-quarters of trials, the effect is reversed, showing up as a significant decrease in quantity of first-error at the second position of the doublet. A more precise determination of the relation between the magnitude of the doublet effect and stage of acquisition is not possible because of marked individual differences in the number of trials required to reach the criterion performance. *It is clear, however, that the occurrence of a doublet in a maze pattern has an initial inhibitory effect upon acquisition during the discovery phase and that this is a function of guessing habits of Ss, particularly the tendency not to repeat a response correct at the previous position. Once discovery is completed, the similarity of the items* of the doublet facilitate further learning of the second item of the doublet. The latter hypothesis is supported only by results for the first-error measures, although the serial position curves constructed from the trial data support this in part by showing evidence of a decrease in trials for the second position of the doublet.

The split-doublet effect does not occur when the locus is early in the maze pattern and its nature for other locations is not completely consistent. It does involve an increase in error at the third position of the split-doublet sequence, and possibly also an increase for the second position of smaller magnitude than for the third position. There is no evidence of a difference in curve form for total errors and first errors, for successive stages of acquisition, nor of a relation between the guessing habits of *S*s and the split-doublet effect.

The results of these experiments suggest that in serial learning, variables of pattern may facilitate acquisition for the positions of their location or may inhibit it. When the serial learning involves discovery, as in the case of verbal mazes, the variables of pattern may or may not interact with the guessing habits of *S*s to provide either for facilitation or inhibition of acquisition. The location of the pattern variable in the sequence may be differential in producing an effect and on the magnitude of the effect. In the case of magnitude, the middle and late positions of the sequence may be more sensitive to the effect of pattern than are the early positions.

Summary

Two experiments were designed to test the effect of location of a doublet (two successive identical responses) and a split-doublet (two identical responses separated by one different response) on the form of the serial position error curve from the acquisition of a 4-choice, 16-unit verbal maze. In both experiments, significant differences were obtained in the form of the curves between each experimental group and the control group, and between experimental groups. In the case of the doublet, for each of the five loca-

tions, the effect was an increase in total error at the second position of the doublet. The effect did not occur for the first error data over all trials, but when the data for the first quarter and last three-quarters of trials were analyzed separately, an inhibitory effect was found for the first quarter and a facilitative effect was found for the last three-quarters of trials. The inhibitory effect is demonstrated to be a function of the guessing tendencies of the *S*s not to repeat a response previously correct during the discovery phase. The split-doublet produced an increase in error for the third position for three of the four locations, and in two of these there was also an increase in error at the second position of the split-doublet. No relation was found between the guessing tendencies of *S*s and the split-doublet phenomena, nor were there any differences in curve form between total errors and first errors or as a function of stage of acquisition.

REFERENCES

BROGDEN, W. J., & SCHMIDT, R. E. Effect of number of choices per unit of a verbal maze on learning and serial position errors. *J. exp. Psychol.*, 1954, **47**, 235–240.

McCRARY, J. W., JR., & HUNTER, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131–134.

THOMPSON, R. F., VOSS, J. F., & BROGDEN, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, 1957, **54**, 253–258.

(Received March 4, 1959)

Journal of Experimental Psychology
Vol. 59, No. 2, 1960

RETROACTIVE INHIBITION OF CONNECTED DISCOURSE AS A FUNCTION OF PRACTICE LEVEL[1]

NORMAN J. SLAMECKA

University of Vermont

Retroactive inhibition (RI) varies inversely as a function of the degree of practice on original learning (OL) (McGeoch, 1929; Briggs, 1957), and, up to intermediate levels of interpolated learning (IL), directly as a function of the degree of practice on IL (McGeoch, 1932; Melton & Irwin, 1940; Underwood, 1945). All of these studies used nonsense syllables or other unconnected materials. However, there is doubt about the susceptibility of connected discourse to RI, particularly when rote retention measures are used (McGeoch & McKinney: 1934a, 1934b). We have been unable to find any evidence in the published literature that verbatim recall of prose is subject to significant RI. However, some preliminary work from this laboratory, using a group testing procedure, gave encouraging results (Slamecka, 1958), and the general importance of this problem warranted further investigation.

The present study was designed as a more precise test to determine whether varying degrees of practice would have the same effect upon the recall of connected discourse as they have upon unconnected materials. Specifically, it was hypothesized that the rote retention of a passage of prose is subject to RI, and that its recall is a direct function of amount of practice in OL and an inverse function of amount of practice in IL.

METHOD

Subjects and materials.—Thirty-six students in general psychology at the University of Vermont served as *S*s as part of the requirements of the course. Each *S* served in four 40-min. sessions, the first being a practice period. Each session followed the RI paradigm, with OL, IL, and relearning of the OL passage by the method of serial anticipation. The materials were eight 20-word sentences, all taken from the same book source, modified slightly for length. Each word was exposed on the memory drum for 3 sec. Thus each trial took 60 sec., followed by a 6-sec. intertrial interval. The passages were grouped into four pairs, one of which was used for practice, and the others for the experiment proper. The first passage of each pair was for OL and the second for IL. The following three pairs of passages were used in the experiment proper:

(*a*) We must postulate that from strictly semantic points of vantage, most confusions in communication revolve about inadequate stipulation of meaning.

Communicators can exercise latitude in specifying meaning however they choose, provided that such definitions correspond somewhat closely to customary usage.

(*b*) So habitual become our expectations about symbols invariably possessing referents, that we tend unconsciously toward such assumptions concerning every word.

When words are utilized ostensibly as symbols although lacking appropriate referents, then perilous depths of semantic ignorance have been reached.

(*c*) However in most instances, linguists are predominantly concerned about determining the concrete denotations of discrete words functioning as proper names.

Much evidence regarding the inference of authentic meanings which are not stipulated, is derived from context analysis within the passage.

---

[1] This research was sponsored by National Science Foundation Grant G-6192. The author is indebted to Ernest N. Damianopoulos who assisted in the collection and analysis of the data.

TABLE 1

Means and *SD*s of Total Correct Anticipations per *S* during OL and IL Acquisition

IL Trials	OL-2		OL-4		OL-8	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
	OL Acquisition					
0	18.00	5.00	40.33	12.88	105.17	21.16
4	17.42	6.38	40.17	8.69	107.17	22.12
8	16.33	5.57	44.50	11.30	97.33	22.24
	IL Acquisition					
4	32.58	8.6	37.08	15.66	43.83	11.24
8	89.00	21.27	109.25	20.22	114.75	26.10

Design and procedure.—A mixed factorial design (Lindquist, 1953, Type II) was used. One variable was the degree of OL practice (2, 4, and 8 trials), and the other was the degree of IL practice (0, 4, and 8 trials). Of the nine possible combinations of these levels of both variables, each *S* was given three, in such a manner that he had all levels of each variable but no repetitions of any level. The 36 *S*s provided 12 replications of the nine-cell matrix. Conditions were arranged so that the list pairs, sequence of sessions, and *S*s, were each counterbalanced with respect to the two main variables.

Each session began with two warm-up trials of color guessing. The colors appeared on the memory drum. Next, instructions for serial anticipation were given, followed by the appropriate number of OL and IL trials. In every case, 30 sec. elapsed between the end of OL and the start of IL, and 9.5 min. elapsed between the completion of OL and the OL relearning. To fill the temporal gap for the control (IL-0) and IL-4 conditions, a "rest activity" of color guessing was used in order to prevent rehearsal, while still preserving the set to respond to stimuli on the drum. Then, four relearning trials on the original passage were given, followed by three trials on the interpolated passage. Relearning of IL was done to motivate *S*s equally in the acquisition of OL and IL, thereby preventing laxness with respect to IL in later sessions.

TABLE 2

Analyses of Variance of OL Acquisition and Recall

Source	*df*	Acquisition		Recall	
		MS	*F*	*MS*	*F*
Bet. *S*s	35	.031	.97	.069	1.07
OL × IL (b)	2	.016	.50	.142	2.20
Error (b)	33	.032		.065	
Within Ss	72	.165	17.85*	.114	2.93*
IL	2	.002	.20	1.088	28.06*
OL	2	5.629	608.67*	1.623	41.85*
OL × IL (w)	2	.006	1.25	.099	2.55
Error (w)	66	.009		.039	

* $P < .01$.

Results and Discussion

Acquisition data for OL and IL appear in Table 1, in terms of the mean total correct anticipations per *S*. The OL data were subject to analysis of variance after a logarithmic transformation, and the results appear in the left half of Table 2. It is clear that increasing the OL trials, at every level of IL, produced a significant increase in the mean correct anticipations. Means within any IL level differed significantly from the adjacent means and there were no significant *F*s within any OL level. Analysis of the IL acquisition data indicated that increasing the

IL practice significantly increased the degree of acquisition. A tendency toward positive transfer was evident in that the IL means increased with increasing OL trials. Since this worked in a direction counter to the hypothesized effects of the main variables, the significance of the recall analysis, to follow, can only tend toward conservatism.

Table 3 shows the relearning of OL in terms of the mean total correct anticipations on each relearning trial. The recall data (first relearning trial) were analyzed after a log $(1 + X)$ transformation, and the relevant findings appear in the right half of Table 2. Both of the main variables were seen to be effective in determining recall. Within the levels used, recall as a function of OL practice approximates a linear function. An extrapolation through the OL-0 point gave a value of about two correct anticipations. It would be expected that even with no practice on such passages, *S* could guess a few words correctly by virtue of previous familiarity with the general language structure. Analyses within each OL level and within each IL level gave significant *F*s. Regardless of the degree of OL practice, the control group (IL-0) always recalled significantly more than the experimental groups (IL-4, IL-8). Thus, significant RI was obtained. It should be noted that the direction of effects of amounts of OL and IL was always as predicted even though all specific comparisons were not significant.

From the data of Table 3, relative RI $\left[\left(\frac{\text{Rest-Work}}{\text{Rest}}\right) \times 100\right]$, was calculated. The over-all RI was 43.0%. At IL-4 the relative RI for OL-2, 4, and 8, respectively, was 58.7%, 38.6%, and 24.3%. At IL-8 the values were 76.0%, 44.1%, and 34.7%, respectively. Thus, relative RI decreased as a function of increasing OL and decreasing IL practice.

The forgetting was rapidly dissipated within a few relearning trials. Table 3 shows that by the fourth relearning the differential effects of the interfering IL practice upon recall were largely dissipated, which

TABLE 3

MEANS AND *SD*s OF OL RELEARNING TRIALS

RL Trials	IL Trials	2 OL Trials		4 OL Trials		8 OL Trials	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
1	0	8.67	3.00	9.08	4.06	14.42	3.80
	4	3.58	3.16	5.58	2.53	10.92	3.10
	8	2.08	1.98	5.08	3.62	9.42	4.90
2	0	11.75	1.79	13.16	4.33	16.83	3.41
	4	10.16	3.53	12.50	2.66	16.25	2.83
	8	9.08	3.82	12.16	4.54	14.33	4.06
3	0	14.75	2.58	14.75	3.74	17.91	3.22
	4	11.25	3.48	14.50	2.30	18.50	1.65
	8	11.83	3.93	14.50	3.75	15.66	3.10
4	0	15.25	3.19	14.75	4.44	18.91	2.10
	4	13.16	5.48	15.66	1.79	19.33	1.18
	8	13.33	4.56	15.58	3.22	17.00	2.47

TABLE 4

CLASSIFICATION OF ABSOLUTE NUMBER OF ERRORS AT RECALL

Error Class	IL-0			IL-4			IL-8		
	OL 2	OL 4	OL 8	OL 2	OL 4	OL 8	OL 2	OL 4	OL 8
Omissions	100	68	37	141	129	73	174	145	92
Intralist	10	32	18	8	10	16	13	12	7
Interlist	0	0	0	7	12	10	5	5	12
Extraneous	24	25	11	32	21	6	23	15	15

is in accord with the usual findings based on unconnected materials.

The types of errors that occur during recall are often scrutinized for hypotheses concerning the processes underlying RI. A tabulation of all responses classed as errors is presented in Table 4. There are four categories of errors: (*a*) omissions, i.e., no response was made; (*b*) intralist intrusions, i.e., incorrectly placed words from the original passage; (*c*) interlist intrusions, i.e., words from the interpolated passage; and (*d*) extraneous intrusions, i.e., words found in neither passage. The use of connected discourse makes it impossible to categorize all intrusions unequivocally. For instance, a response which is classed as an intralist error is still a word in *S*'s language, and may have really originated as an extraneous intrusion. In addition, since connected discourse is serial in nature, errors cannot be subject to the S-R intrusion analysis traditionally done with the A-B, A-C paired-associate paradigm. In spite of these inherent limitations the data appear to provide tentative hints regarding the distribution of response tendencies.

The error data support two generalizations. First, that the better the recall, the less likely is an error to be an omission. This is the tendency even when omissions are stated as percentages of all errors made. Thus, under conditions of maximum recall (OL-8 and IL-0), only 56% of the errors are omissions, whereas under conditions of minimum recall (OL-2 and IL-8), 81% of all errors are so classed. Considering the two variables singly, there is still a trend toward decrease in the percentage of omissions as OL increases, and also as IL decreases. If omissions are considered as covert errors, and all other errors are considered as overt errors, then the data suggest a shift in the covert–overt error ratio paralleling changes in recall, reminiscent of the formulation of Thune and Underwood (1943).

The second generalization concerns the appearance of interlist intrusions. The frequency of these errors has been used as the index to the degree of competition attributed to the interfering material. The data suggest that such intrusions (whether listed in absolute or percentage terms) are more frequent when OL and IL acquisition is about equal. In terms of the percentage of all overt errors accounted for by this category, higher values are found at the OL-8-IL-8 (35%), and the OL-4-IL-4 (28%) levels, where the two lists are at about equal strength. Such a tendency has been reported for unconnected materials by Briggs (1957) and others. It appears to be consistent with a differentiation hypothesis (Gibson, 1941) whereby competing responses of near equal strength are most likely to produce such intrusions, as compared to response systems more differentiated from each other by unequal amounts of practice.

With respect to the two hypotheses tested in this study, the following conclusions are drawn. First, that the rote retention of connected discourse has been shown to be subject to significant RI, and secondly, that such recall varies directly with degree of OL practice and inversely with degree of IL practice. These results serve to confirm the previous preliminary findings (Slamecka, 1958) and support the extension of generalizations hitherto based on the serial learning of unconnected materials to the serial learning of connected discourse.

SUMMARY

In the RI paradigm, *S*s learned passages of connected discourse by the anticipation method. Three levels of OL practice (2, 4, and 8 trials) and three levels of IL practice (0, 4, and 8 trials) were given, with appropriate counterbalancing in a mixed design. Results indicated that (*a*) OL and IL acquisition was a function of number of practice trials, (*b*) recall of prose was subject to significant RI, (*c*) recall varied directly with OL practice and inversely with IL practice, and (*d*) recall errors suggested a shift in the covert–overt error ratio paralleling recall, and an increase in interlist intrusions at points of near equal acquisition of the two passages. It is concluded that findings regarding degree of learning and RI, based on unconnected materials, can now be generalized to connected discourse.

REFERENCES

BRIGGS, G. E. Retroactive inhibition as a function of the degree of original and interpolated learning. *J. exp. Psychol.*, 1957, **53**, 60–67.

GIBSON, E. J. Retroactive inhibition as a function of the degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCGEOCH, J. A. The influence of degree of learning upon retroactive inhibition. *Amer. J. Psychol.*, 1929, **41**, 252–262.

MCGEOCH, J. A. The influence of degree of interpolated learning upon retroactive inhibition. *Amer. J. Psychol.*, 1932, **44**, 695–708.

MCGEOCH, J. A., & MCKINNEY, F. Retroactive inhibition in learning of poetry. *Amer. J. Psychol.*, 1934, **46**, 19–33. (a)

MCGEOCH, J. A., & MCKINNEY, F. The susceptibility of prose to retroactive inhibition. *Amer. J. Psychol.*, 1934, **46**, 429–436. (b)

MELTON, A. W., & IRWIN, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

SLAMECKA, N. J. Retention of connected discourse as a function of degree of learning. Paper presented at APA meetings, Washington, September, 1958.

THUNE, L. E., & UNDERWOOD, B. J. Retroactive inhibition as a function of degree of interpolated learning. *J. exp. Psychol.*, 1943, **32**, 185–200.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, (**59** 3, Whole No. 273).

(Received March 7, 1959)

Journal of Experimental Psychology
Vol. 59, No. 2, 1960

MOTOR SKILL TRANSFER AS A FUNCTION OF INTERTASK INTERVAL AND PRETRANSFER TASK DIFFICULTY

GEDIMINAS NAMIKAS AND E. JAMES ARCHER

University of Wisconsin

In a recent pursuit-rotor study Lordahl and Archer (1958) examined the effects of transferring from one speed of rotation to a second speed of rotation. The *S*s practiced tracking a rotary-pursuit target at 40, 60, or 80 rpm on Day 1 and on Day 2 all *S*s practiced at 60 rpm. Substantial negative transfer obtained when there was a change in target speed.

The present study is concerned with the effect of variation of intertask interval upon this negative transfer. It was suggested in the Lordahl and Archer paper that, among other things, *S* learned a certain rhythm of responding. Conceivably the recovery from or forgetting of this rhythm could vary as a function of time since practice.

PROCEDURE

Subjects.—Women students from introductory psychology courses at the University of Wisconsin, who volunteered for the experiment, were used as *S*s. These *S*s received extra credit toward their course grade for serving in the experiment. Ten *S*s were assigned randomly to one of 12 conditions ($N = 120$). Four extra *S*s were run and their data discarded due to failure of either the tracking or recording apparatus. All *S*s were naive to the task and apparatus.

Apparatus and task.—The *S*s tracked a $\frac{3}{4}$-in. diameter, flush-mounted, brass target which was set 5 in. from the center of a pressed wood turntable. The turntable had a radius of 6 in. The tracking surface was 35 in. from the floor. The *S* tracked the target with a hinged stylus which prevented *S* from pressing on the handle and holding the stylus on target by force. The length of the stylus from the hinge to the point was 6 in. and the tip contained a steel ballbearing as illustrated in a previous study by Archer (1958).

The speed of the turntable was adjustable from about 10 to 85 rpm and a stroboscopic attachment allowed *E* to adjust the speed to within .5 rpm of the desired speeds. Performance of *S* was recorded on the Tracking Time Analyzer (TTA) which has been described in a previous paper (Archer, 1958).

Verbal signals of "Get set" and "Start" and "Stop" were given by *E* at the start and end of each trial respectively. The TTA and *E* were separated from *S* by a partial wall and *S*'s performance was not made known to her. The TTA was housed in a partially soundproof box, and a large noisy electric fan completed the masking of the clicking of its relays. The TTA was calibrated twice a day—morning and afternoon—to correct for fluctuations in line voltage.

Conditions.—All *S*s were given 20 training and 20 transfer trials. The trials were 30 sec. long with 60-sec. rest intervals interpolated between each. The *S*s trained for 20 trials on either the 40, 60, or 80 rpm speed and then had a rest interval of 2, 4, 8, or 16 min. All *S*s then transferred to the 60-rpm speed for 20 trials. Thus a 3 × 4 factorial design with speed of rotation orthogonal to intertask interval was used. During the intertask interval *E* changed the speed of rotation in the case of the 40- and 80-rpm groups and pretended to do so for the 60-rpm group by first changing the speed of rotation and then returning to the same speed. The three groups will hereafter be designated 40–60, 60–60, and 80–60.

Instructions.—All *S*s were given general instructions about the best way to track the target and a demonstration of such tracking by *E*. They were told to stay on target as long as possible. The *S*s were not told the number of trials they were going to have but the length of the trial and the rest interval were indicated. They were not informed about the longer rest interval, i.e., the intertask interval, beforehand, but after the training trials *E* told each *S* that the speed of rotation was going to be changed and that they were going to have a rest of a given duration. They were instructed to pick up the handle of the stylus on the signal "Get set" and then wait until the "Start" signal was given before starting to track the target.

During the intertrial and intertask intervals, after the signal "Stop," *S*s were told to place the tip of the stylus in the center of the turntable and to sit down and look at magazines which were placed near by.

The *S*s were not told anything about the recording apparatus. The answers to any questions, except those relating to the instructions, were deferred until the end of the experimental session. Instructions were tape recorded and were therefore identical for all *S*s.

RESULTS

The results are summarized in Fig. 1. This figure shows the mean percentage of time on target for all speed-interval combinations for both the training and transfer sessions.

Training session.—An extended Alexander trend analysis of variance (Grant, 1957) was performed on the training data. The intertask interval classification was included as a dummy variable to test the adequacy of the random assignment procedure and the assumption of initially equal subgroups within each speed condition. Since the test indicated that intertask interval was not an effective variable the assumption of equal subgroups seemed tenable.

The same analysis indicated that the speed of rotation was a highly effective variable. This agrees with the results of the Lordahl and Archer (1958) study. Time on target was found to be an inverse function of speed of rotation and the between-slopes and quadratics terms were highly significant (*F*s = 221.63 and 63.44, respectively; $P < .01$ with 2 and 108 *df*). The significant linear component indicates that performance increases differentially as a function of

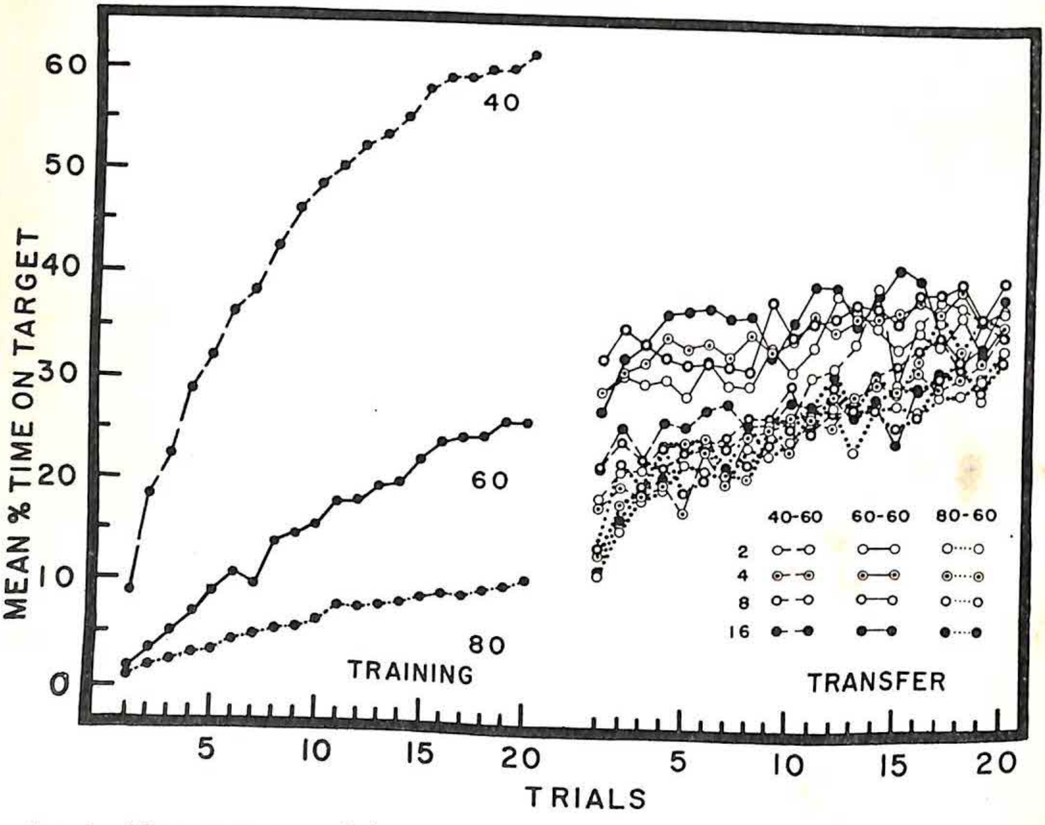


FIG. 1. Mean percentage of time on target for the training and transfer stages. Each data point in the training curves is based upon 40 cases; in the transfer stage each point is based upon 10 cases.

TABLE 1

A FREQUENCY DISTRIBUTION OF TIME CONTINUOUSLY ON TARGET IN THE FIRST FIVE TRIALS OF THE TRANSFER TASK

Training Task	Time Continuously On Target (In Sec.)										Total
	.00–.04	.05–.09	.10–.19	.20–.29	.30–.39	.40–.59	.60–.79	.80–.99	1.00–1.19	1.20–1.59	
40 rpm	2497	1525	2173	854	500	449	88	28	15	14	8143
60 rpm	2867	1849	2852	1137	792	692	230	73	50	33	10,575
80 rpm	3014	1600	1920	674	449	327	79	28	19	6	8116

Note.—The data are the total numbers of hits within each of the duration categories, with $N = 40$ for each training—task condition.

the speed of rotation, while the quadratic term shows a differential rate of increase in performance as a function of speed.

Transfer session.—An extended Alexander trend analysis of variance was performed on the transfer data also. Speed of rotation, during the training task, seems to have been the only variable to affect the transfer performance. There were significant differences between the means, linear, and quadratic components for speed ($F = 7.774$, $F = 13.473$, $P < .01$; $F = 3.550$, $P < .05$ with 2 and 108 *df*).

Because the total range of variation of the independent variable of intertask interval was relatively short (16 min.), it is conceivable that the influence of the intertask interval would be evident in only the first few trials on the transfer task. Accordingly, an Alexander trend analysis of variance was performed on the time-on-target data for the first five transfer trials only. Again, only the speed of the training task proved to have any significant effect. As in the analysis of the 20 transfer trials the between-group means, slopes, and quadratic components were found to be significant.

An extended trend analysis of variance was performed on the first five trials after transfer on the data from just the 40–60 and the 80–60 groups. No significant differences were found indicating that the above results were due to the differential performance of the 60–60 group as compared to the 40–60 and 80–60 groups.

Distribution of hits.—In Table 1 is presented the distribution of hits (frequency of times continuously on target) for the first five transfer trials. The numbers of hits for each duration were combined within each speed group without regard to the intertask interval. The superior performance of the 60–60 group is evident when the number of long-duration hits scored by this group is compared with the number scored by each of the other two groups (40–60 and 80–60). A chi square of 398.58 for 18 *df* ($P < .001$) was found for this table. These results are essentially the same as found by Lordahl and Archer (1958) and indicate that about the same transfer effects are found after an average of about 8 min. rest as compared to the 24-hr. interval studied by Lordahl and Archer. In order to complete both training and transfer sessions within a reasonable experimental period both of the sessions in the present study were made somewhat shorter than those in the Lordahl and Archer study. Therefore the magnitude of negative transfer after 8 min. and 24 hr. is not directly comparable.

DISCUSSION

Speed of rotation.—The results substantiate the previous findings that variation in task difficulty, in terms of speed of rotation, has a very marked effect upon performance on both the training task and the transfer task.

In training, performance is found to vary inversely with the speed of rotation, i.e., better performance is found with slower speeds. The training condition was also found to have a very significant effect upon transfer performance. A major effect is due to a *change* in tracking speed. Since even in the first five trials after transfer, where the differences seemed to be greatest, there were no significant differences between the 40–60 and 80–60 groups, the above differences must have been due entirely to the better performance of the 60–60 group.

As suggested in the Lordahl and Archer study (1958) a possible hypothesis to explain these findings is that *S* learns a given set of motor components for performing a particular task. It is probable that a task of a given degree of difficulty emphasizes a particular set of components, which are acquired by *S*. With a change in difficulty, both training and transfer performance would be affected. For instance, an *S* in the 40-rpm group would probably learn a set of smooth, fairly slow, circular motions. These would tend to produce long duration hits and large time-on-target scores if he ever got on target. This *S* could be compared with one training in an 80-rpm condition. In this case a rapid jerky motion emphasizing quick adjustment to being off target would be acquired. Such a condition should produce a greater number of hits but most of them of short duration and probably with a smaller time on target score. On transferring to an intermediate difficulty condition, both of these *S*s would be under a handicap until they adjusted their performance to emphasize a particular combination of components most appropriate to this new condition. The rapidity of this adjustment would probably vary with the perceived similarity of the two tasks and the ability of *S* to realize the inappropriateness or inefficiency of his response.

Intertask interval.—Although *S* may learn the above hypothesized component skills appropriate to the training task difficulty, the performance of these component skills does not seem to vary differentially over a period of 16 min. In fact, the amount of negative transfer for any one group does not appear to vary as a function of these relatively short intertask intervals. Results indicate that the particular lengths of intertask intervals used in this study were completely ineffective in producing concomitant variations in performance.

SUMMARY

The effects of three conditions of task difficulty were investigated on training and transfer performance on a rotary-pursuit task. The transfer was to the intermediate condition of difficulty used in training. The difficulty was varied by varying the speed of rotation in a rotary-pursuit task. Three speeds (40, 60, and 80 rpm) were used in training. All *S*s switched to 60 rpm in the transfer task.

Performance on transferring was also investigated as a function of the length of intertask interval. Four intertask intervals were used (2, 4, 8, and 16 min.).

The main findings were: (*a*) task difficulty (speed of rotation) was a very effective variable in both training and transfer performance, (*b*) performance was inversely related to task difficulty under training conditions, (*c*) transferring to the same speed gave the best performance, and (*d*) intertask interval did not affect transfer performance.

REFERENCES

ARCHER, E. J. Effect of distribution of practice on a component skill of rotary pursuit tracking. *J. exp. Psychol.*, 1958, **56**, 427–436.

GRANT, D. A. Analysis of variance in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

LORDAHL, D. S., & ARCHER, E. J. Transfer effects on a rotary pursuit task as a function of first-task difficulty. *J. exp. Psychol.*, 1958, **56**, 421–426.

(Received March 9, 1959)

Journal of Experimental Psychology
Vol. 59, No. 2, 1960

ANCHOR, CONTRAST, AND PARADOXICAL DISTANCE EFFECTS

HARRY HELSON
University of Texas

AND MYRTLE C. NASH
Southwestern at Memphis

The paradoxical distance effect (PDE) reported by Köhler and Wallach (1944) refers to the fact that test figures exposed in the vicinity of satiated or adapted visual areas suffer greater displacement the greater their distance, within limits, from the originally fixated figure. This finding is paradoxical because it appears to be in conflict not only with the almost universal physical law according to which effects diminish with distance from their source, but also with such basic psychological phenomena as lightness contrast and anchor effects where the rule is that stimuli near backgrounds or anchors suffer greater displacement than do stimuli farther away. While the theory of electronic currents advanced by Köhler and Wallach for figural aftereffects presumably accounts for PDE, it nevertheless leaves this interesting and important phenomenon in a class by itself. On the one hand it is reasonable to assume that lightness contrast and PDE are related since both result from interaction of retinal and/or central visual areas, and on the other it is not unreasonable to assume that phenomena in other sense modalities which are either materially or formally similar to PDE may be explicable in terms of a model which is independent of sense modality and which also accounts for background and anchor effects. While investigating effects of anchors on sets of weights at various distances from the anchoring stimuli, two types of shifts in judgments were found, one of which is analogous to classical contrast effects and the other to PDE. This study is concerned with these results and the possibility of dealing with them by means of a single theoretical model.

METHOD

Stimuli and procedure.—The stimuli consisted of 4-oz. medicine bottles filled with sand or mercury to make seven sets of weights as follows: Set I: 100, 150, 200, 250, and 300 gm.; Set II: 150, 200, 250, 300, and 350 gm.; Set III: 200, 250, 300, 350, and 400 gm.; Set IV: 250, 300, 350, 400, and 450 gm.; Set V: 300, 350, 400, 450, and 500 gm.; Set VI: 350, 400, 450, 500, 550 gm.; Set VII: 400, 450, 500, 550, and 600 gm. The sets of stimuli were presented either in ascending order, beginning with Set I and ending with Set VII, or in descending order, beginning with Set VII and ending with Set I. The weights in each set were judged in random order, and after five presentations the next set was judged without any break. To minimize interset effects, a control procedure was employed in which *Ss* judged Set VII immediately after Set I or Set I immediately after Set VII. Furthermore, the weights were first judged singly and one week or two weeks later were judged by all *Ss* following either a 90-gm. or a 900-gm. background stimulus. Counting the two orders of presentation (ascending and descending),

TABLE 1

Judgments of Seven Sets of Weights in Ascending and Descending Orders by Method of Single Stimuli (S.S.) and Following a 900-Gm. Background Stimulus (B.S.) Showing Differential Effects of the Background on Stimuli within Sets (Contrast) and between Sets (Paradoxical Distance Effect)

Stimuli:	100	150	200	250	300	350	400	450	500	550	600	Total
Ascending Series Order												
S.S.:	20	41	61	75	86							
900 B.S.:	7	16	27	39	49							
Diff.:	−13	−25	−34	−36	−37							−145
S.S.:		34	48	62	72	80						
900 B.S.:		12	24	31	41	50						
Diff.:		−22	−24	−31	−31	−30						−138
S.S.:			39	58	70	83	89					
900 B.S.:			17	27	35	44	49					
Diff.:			−22	−31	−35	−39	−40					−167
S.S.:				48	64	72	81	90				
900 B.S.:				29	38	44	52	60				
Diff.:				−19	−26	−28	−29	−30				−132
S.S.:					50	62	76	82	89			
900 B.S.:					35	43	51	56	62			
Diff.:					−15	−19	−25	−26	−27			−112
S.S.:						55	65	77	86	88		
900 B.S.:						41	45	55	62	68		
Diff.:						−14	−20	−22	−24	−20		−100
S.S.:							58	68	80	83	88	
900 B.S.:							47	57	64	69	73	
Diff.:							−11	−11	−16	−14	−15	−67
Descending Series Order												
S.S.:							44	56	69	77	80	
900 B.S.:							41	49	57	64	71	
Diff.:							−3	−7	−12	−13	−9	−44
S.S.:						40	54	66	74	83		
900 B.S.:						33	40	46	56	63		
Diff.:						−7	−14	−20	−18	−20		−79
S.S.:					33	47	60	71	80			
900 B.S.:					26	33	40	49	56			
Diff.:					−7	−14	−20	−22	−24			−87
S.S.:				30	45	56	70	80				
900 B.S.:				19	26	37	45	52				
Diff.:				−11	−19	−19	−25	−28				−102
S.S.:			27	43	56	69	80					
900 B.S.:			18	23	30	38	47					
Diff.:			−9	−20	−26	−31	−33					−119
S.S.:		23	37	51	68	77						
900 B.S.:		13	21	30	36	45						
Diff.:		−10	−16	−21	−32	−32						−111
S.S.:	22	36	51	66	75							
900 B.S.:	9	16	22	34	43							
Diff.:	−13	−20	−29	−32	−32							−126

TABLE 2

JUDGMENTS OF SEVEN SETS OF WEIGHTS IN ASCENDING AND DESCENDING ORDERS BY METHOD OF SINGLE STIMULI (S.S.) AND FOLLOWING A 90-GM. BACKGROUND STIMULUS (B.S.) SHOWING DIFFERENTIAL EFFECTS OF THE ANCHOR ON STIMULI WITHIN SETS (CONTRAST) AND BETWEEN SETS (PARADOXICAL DISTANCE EFFECT)

Stimuli:	100	150	200	250	300	350	400	450	500	550	600	Total
Ascending Series Order												
S.S.:	20	37	54	70	75							+33
90 B.S.:	36	47	60	68	78							
Diff.:	+16	+10	+6	−2	+3							
S.S.:		35	50	63	72	82						+37
90 B.S.:		48	60	69	78	84						
Diff.:		+13	+10	+6	+6	+2						
S.S.:			46	56	68	78	83					+48
90 B.S.:			62	70	78	82	87					
Diff.:			+16	+14	+10	+4	+4					
S.S.:				48	60	70	76	84				+51
90 B.S.:				67	72	79	84	87				
Diff.:				+19	+12	+9	+8	+3				
S.S.:					57	65	71	78	84			+48
90 B.S.:					69	77	82	86	89			
Diff.:					+12	+12	+11	+8	+5			
S.S.:						60	68	76	80	86		+53
90 B.S.:						77	81	86	88	91		
Diff.:						+17	+13	+10	+8	+5		
S.S.:							65	73	79	82	88	+45
90 B.S.:							78	82	91	91	90	
Diff.:							+13	+9	+12	+9	+2	
Descending Series Order												
S.S.:							39	50	61	74	77	+73
90 B.S.:							65	69	77	80	83	
Diff.:							+26	+19	+16	+6	+6	
S.S.:						37	47	57	70	75		+87
90 B.S.:						62	70	79	80	82		
Diff.:						+25	+23	+22	+10	+7		
S.S.:					34	46	55	65	70			+76
90 B.S.:					57	62	69	78	80			
Diff.:					+23	+16	+14	+13	+10			
S.S.:				32	43	51	64	74				+57
90 B.S.:				50	56	66	72	77				
Diff.:				+18	+13	+15	+8	+3				
S.S.:			26	39	54	62	71					+48
90 B.S.:			43	53	59	69	76					
Diff.:			+17	+14	+5	+7	+5					
S.S.:		24	35	50	60	68						+27
90 B.S.:		33	44	56	62	69						
Diff.:		+9	+9	+6	+2	+1						
S.S.:	16	29	45	54	64							+16
90 B.S.:	24	34	46	54	66							
Diff.:	+8	+5	+1	0	+2							

TABLE 3

SHIFTS IN ADAPTATION-LEVEL IN DIRECTION OF BACKGROUND STIMULI WITH GREATER MOVEMENT IN SETS FARTHER FROM VALUE OF BACKGROUND AS ANALOGUE OF PDE (ASCENDING ORDER)

Stimulus Sets	Adaptation-Level		Diff.	Adaptation-Level		Diff.
	Single Stimuli	900-Gm. Background		Single Stimuli	90-Gm. Background	
100–300	165	317	+152**	181	155	−26
150–350	202	357	+155**	198	159	−39
200–400	227	404	+177**	219	137	−82**
250–450	257	380	+123**	258	144	−114**
300–500	301	398	+97**	270	173	−97**
350–550	332	415	+83**	300	119	−181**
400–600	361	409	+48*	312	228	−84**

* $P < .05$.
** $P < .01$.

the two background stimuli, and the two modes of transition from lightest (heaviest) to heaviest (lightest) sets of stimuli (gradual and abrupt), it is seen that there were eight different conditions of judging. Different groups of 5 *S*s each judged under each of the eight conditions, thus making a total of 40 *S*s who served in this experiment.

Instructions.—The *S*s were instructed as follows: "You will be presented with weights in succession. After lifting each weight you are to give a judgment by calling it very, very heavy; very heavy; heavy; medium heavy, medium, medium light; light; very light; or very, very light. If you need other categories such as extremely heavy or extremely light or terrifically light or terrifically heavy, you may use them. Be sure to judge each weight as it is and not as you think it should be." When the background weights were introduced, *S*s were instructed to judge the second weight by itself, not as a comparison with the first weight. To acquaint *S*s with stimuli and with the category scale, two practice rounds were given with the set of weights first presented.

Treatment of data.—The judgments given by *S*s in terms of the rating scale were translated into numbers ranging from 10 for very, very light to 90 for very, very heavy in steps

TABLE 4

SHIFTS IN ADAPTATION-LEVEL IN DIRECTION OF BACKGROUND STIMULI WITH GREATER MOVEMENT IN SETS FARTHER FROM VALUE OF BACKGROUND AS ANALOGUE OF PDE (DESCENDING ORDER)

Stimulus Sets	Adaptation-Level		Diff.	Adaptation-Level		Diff.
	Single Stimuli	900-Gm. Background		Single Stimuli	90-Gm. Background	
400–600	419	452	+33	443	294	−149**
350–550	384	458	+74*	406	261	−145**
300–500	356	460	+104**	369	268	−101**
250–450	314	437	+123**	327	267	−60*
200–400	269	442	+173**	287	244	−43
150–350	232	404	+172**	251	226	−25
100–300	186	371	+185**	234	215	−19

* $P < .05$.
** $P < .01$.

of 10, with 50 for the medium category. In a very few cases some *S*s employed categories more extreme than very, very light and very, very heavy, and these were assigned numerical ratings in accordance with their judged distance from the center of the rating scale. Values of adaptation-level as well as judgments of the stimuli expected according to theory were found from least squares fits to an equation formulated by Helson (1948) and also Helson and Himelstein (1955). The original ratings of the stimuli (in numerical terms), rather than theoretical values, are given in Tables 1 and 2, but the values of adaptation-level in Tables 3 and 4 are the fitted values. The influence of background (satiating) stimuli on judgments of the series stimuli appears in the differences between judgments of the stimuli presented singly and judgments following the background stimuli and also in the differences between the adaptation-levels for the sets of stimuli as judged in the two conditions. While the background effect is contaminated by carry-over effects from set to set, interset effects are also present when the stimuli are judged singly, and since the latter judgments serve as the norms by which shifts in judgment are gauged when anchors are introduced, interset influences may be regarded as common to both conditions in the comparison. The abrupt transitions from first to last sets and from last to first sets minimize interset effects and furnish a check on the results with continuous transitions from set to set.

RESULTS

Two types of displacement.—The data presented in Fig. 1 and 2 and in Tables 1 and 2 reveal two distinct effects on judgments of series stimuli following introduction of the background stimuli. The first effect is the usual contrast effect in which stimuli nearer background suffer greater displacement than do stimuli farther from background. Thus the plots in Fig. 1 show that the heavier stimuli in both the heaviest and lightest sets of stimuli are displaced

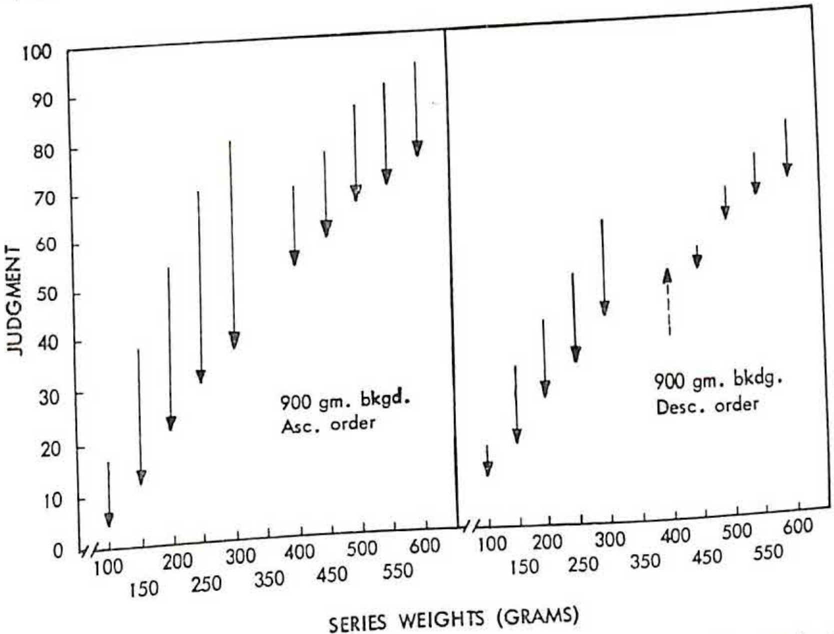


FIG. 1. Shifts in judgments of series stimuli following introduction of heavy background stimulus. The vectors represent direction and amount of shift and they reveal two types of displacements: (*a*) stimuli nearer background stimulus are displaced more than are stimuli farther from background (classical contrast and repulsion effect); (*b*) over-all displacements are greater in sets of stimuli farther from background stimulus (PDE).

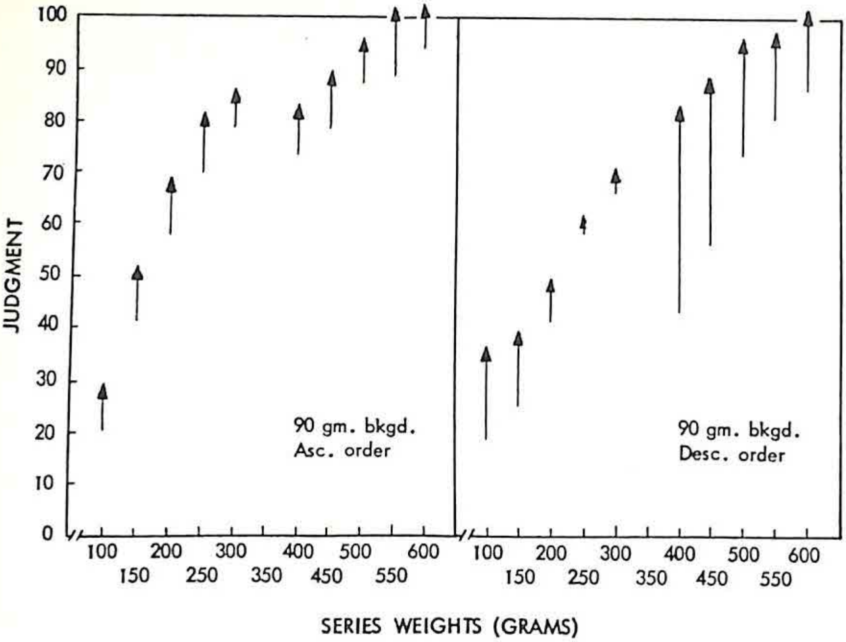


FIG. 2. Shifts in judgments of series stimuli following introduction of light background stimulus. Here the shifts are opposite in direction to those found with heavy background but the two types of displacements are nevertheless present.

more with the 900-gm. background than are the lighter stimuli; the plots in Fig. 2 show that the lighter stimuli within both the lightest and heaviest sets of stimuli are displaced more with the 90-gm. background than are the heavier stimuli. The data in Tables 1 and 2 bear out this generalization for all seven sets of stimuli with only one or two exceptions. Considering the number of *S*s and the variety of stimuli, the data are remarkably consistent in showing displacements similar to classical contrast effects.

The second type of shift in judgment appears when one compares the results of sets with sets either by reference to the total differences between judgments by method of single stimuli and with anchor stimuli (Tables 1 and 2), or by reference to the adaptation-levels in the two methods of judging (Tables 3 and 4). The data in Tables 1 and 2 show that the farther away any given set is from the background stimuli the greater is the total shift in judgments of series stimuli due to introduction of anchors. This effect is exactly analogous to the Köhler-Wallach PDE. Reference to Table 1 shows that the total shift in judgment with the 900-gm. background stimulus is, in general, greater with the lighter sets of stimuli than with the heavier sets of stimuli in both the ascending and descending orders while the data in Table 2 show the opposite occurs with the 90-gm. background stimulus: here the total changes in judgment are greater with the heavier sets of stimuli than with the lighter sets.

The changes in judgments of the series stimuli due to the two anchors or backgrounds reflect differences in the adaptation-levels as seen from Tables 3 and 4. Thus the adaptation-level for the 100–300-gm. set changes from 165 gm. to 317 gm. when the 900-gm. anchor is introduced, a change of 152 gm., while the adaptation-level for the 400–600-gm. set changes from 361 to 409 gm. with the 900-gm. anchor, a change of only 48 gm. On the other hand the introduction of the 90-gm. anchor results in a change of only 26 gm. in the 100–300-gm. set but results in a change of 84 gm. in the 400–600-gm. set. The displacements of the neutral levels are not only greater, with sets farther from the background stimuli, they are, on the whole, also more significant statistically as shown by the *P* values associated with the differences between adaptation-levels for each of the sets of stimuli. The displacements analogous to PDE are also found in the abrupt transition from lightest (heaviest) to heaviest (lightest) sets of stimuli where inter-set effects are minimal so they cannot be attributed to prior stimulation.

DISCUSSION

In the experiments reported here, two effects have been found: classical contrast effects and PDE, the former appearing when judgments of stimuli *within* sets are compared, the latter when judgments from set to set are compared. While the effects found here concern changes in felt weight as a result of introducing anchors much heavier or much lighter than the series weights, they are otherwise comparable to visual color and spatial contrasts and to PDE found in the spatial displacements of vision. Tactile-kinesthetic phenomena have been recognized by many writers, among them Gibson (1933), Köhler and Dinnerstein (1947), and Wertheimer and Leventhal (1958), as falling under the domain of satiation phenomena. Jaffe (1954) has demonstrated that visual extent may influence perceived tactile extent. The visual analogues to our findings are as follows: dark stimuli are lightened relatively more than are light stimuli on dark ground and light stimuli are darkened more than are dark stimuli on light ground (classical contrast), but all stimuli are lightened more on dark ground than they are on light ground while all stimuli are darkened more on light ground than they are on dark ground (PDE). Better still, if sets of grays are viewed on a dark gray ground and on a light gray ground simultaneously, it will be seen that the lighter grays on dark ground cannot be matched by any grays on light ground and the darker grays on light ground cannot be matched by any grays on the dark ground. Hence classical lightness contrast and a lightness PDE can both be observed if sets of gray stimuli are viewed on dark and light backgrounds simultaneously.

PDE depends on carry-over of a previous state of adaptation into the present. The old and new adaptation-levels form a new, intermediate level against which both old and new stimuli are judged. Köhler enunciated what is essentially our point of view when he said: "If there are currents which go with I-objects, there must also be currents corresponding to T-objects. . . . To a degree, we can tell how the obstructions established by the currents of given I-objects must distort the currents of T-objects" (1951, p. 212). In other words, as soon as the test figure appears, the satiated or adapted region must shift in the direction of the test figure and thus a new reference region is established somewhere between the region previously stimulated and the newly stimulated region. It immediately follows that test stimuli farther from the originally satiated region will be displaced more than nearer stimuli provided there is interaction between the old and new levels.

If this analysis, which covers both

classical contrast effects and PDE, is correct, then the adaptation-levels resulting from interaction of series and background stimuli should be displaced more the more the anchor stimuli differ from the mean of the series stimuli. This deduction follows from the definition of adaptation-level as a weighted geometric mean of series, background, and residual stimuli (Helson: 1947, 1948). It is immediately apparent that averaging 900 gm. with the 100–300-gm. series raises the level of this series more than averaging 900 gm. with the 400–600 series raises its level; conversely, the level of the 400–600-gm. series is lowered more than is the level of the 100–300-gm. series by averaging with the 90-gm. background stimulus. The values of adaptation-level calculated from the data of this study agree with predictions from the weighted mean definition of level regarding both direction and extent of change arising from introduction of background stimuli. A single theory therefore appears to be adequate to account for anchor or background and associated "contrast" effects as well as for the spatial displacements found in figural aftereffects.

There is, however, one respect in which we must limit the operation of the weighted mean definition of level. This definition applies only to the extent that there is interaction among stimuli and residual effects of stimulation and if interaction effects grow weaker, say with distance or some other factor, then account must be taken of this fact. Thus PDE is found only within certain limits, for if the test figure is too far from the region of previous stimulation there is no interaction and figural aftereffects may be diminished or even absent. A formula defining spatial levels would therefore require a distance function which would have to be determined for specific conditions of observation. A number of possibilities immediately present themselves and could be worked out within the general framework of the theory proposed here.

SUMMARY

Eight groups of five *S*s each judged seven sets of weights first by method of single stimuli and second after intervals of one or two weeks following either a very light or very heavy background or anchor stimulus. Two types of shifts in judgment due to introduction of the anchor stimuli were found: (*a*) within each of the sets, stimuli nearer the background stimuli were affected more than were stimuli farther from the background, an effect analogous to classical lightness contrast and the usual repulsion effects of anchor stimuli; (*b*) over-all shifts in judgment from set to set were greater in sets farther from than in sets nearer the background stimuli, an effect analogous to the paradoxical distance effect found in satiation experiments. It is shown that both classical contrast effects and paradoxical distance effects may be accounted for in terms of the adaptation-levels arising from interactions between series and background stimuli and between residuals from previous stimulation and present stimulation. Both the weighted mean definition of adaptation-level and the values of adaptation-level derived from the data of this experiment agree in predicting the direction and amounts of shift of judgment due to introduction of background stimuli. A single model would therefore seem to be applicable to classical contrast effects, repulsion effects of anchors, and paradoxical distance effects.

REFERENCES

GIBSON, J. J. Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.

HELSON, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HELSON, H., & HIMELSTEIN, P. A short method for calculating the adaptation-level for absolute and comparative rating judgments. *Amer. J. Psychol.*, 1955, **68**, 631–637.

JAFFE, R. The influence of visual stimulation on kinesthetic figural after-effects. *Amer. J. Psychol.*, 1954, **67**, 668–676.

KÖHLER, W. Relational determination in perception. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951. Pp. 200–243.

Köhler, W., & Dinnerstein, D. Figural after-effects in kinesthesis. In *Miscellanea Psychologica, Albert Michotte.* Louvain Inst. Sup. Philos., 1947. Pp. 196–220.

Köhler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Philos. Soc.*, 1944, **88**, 269–357.

Wertheimer, M., & Leventhal, C. M. "Permanent" satiation phenomena with kinesthetic figural aftereffects. *J. exp. Psychol.*, 1958, **55**, 255–258.

(Received March 16, 1959)

Journal of Experimental Psychology
Vol. 59, No. 2, 1960

ELIMINATING A CONDITIONED GSR BY THE REDUCTION OF EXPERIMENTAL ANXIETY

ROBERT E. SILVERMAN

New York University

The GSR is especially sensitive to aversive motivational events. Experimental anxiety situations, in which aversive stimuli are presented, cause ordinarily neutral stimuli to elicit strong GSRs. This effect is generally termed sensitization or pseudoconditioning, because it does not depend upon the systematic pairing of a neutral stimulus with an aversive stimulus, as is the case in conditioning. Not only the direct presentation of aversive stimuli can produce sensitization, but also the induction of experimental anxiety by the threat of aversive events can do so. For example, Mowrer (1938) has shown that a neutral stimulus elicited GSRs when *S*s were fitted with electrodes, but that the same stimulus did not elicit GSRs when the electrodes were absent.

Since a sensitized response depends upon the presence or the threat of aversive stimuli, we would expect instructions which remove the threat to affect a sensitized GSR in much the same way as Mowrer's removal of the electrodes. However, a conditioned GSR might behave differently because conditioning is dependent, in part, upon the frequent pairings of the neutral and aversive stimulus. As a result of these pairings, the neutral or conditioned stimulus (CS) evokes a response which for a time is not dependent upon the presence of the aversive or unconditioned stimulus (UCS). The CS appears to act as a substitute for the UCS. If this is the case, then instructions about the absence of the UCS should be less effective with a conditioned response than they are with a sensitized response which depends upon the original UCS.

The purpose of the present study was to compare the effects of anxiety-lessening instructions on a conditioned and on a sensitized GSR. Anxiety-lessening instructions are defined as instructions which terminate the threat of aversive stimuli under the control of *E*. The results of two earlier studies indicated that a conditioned GSR (Cook & Harris, 1937) and a conditioned heart rate response (Notterman, Schoenfeld, & Bersh, 1952) can be influenced by instructions. However, in both of these studies the absence of a pseudoconditioning control group and the relatively long interstimulus intervals used make it uncertain whether the instructions were affecting a conditioned response or a sensitized response.

In order to determine whether anxiety-lessening instructions can affect a conditioned GSR as well as a sensitized GSR, three acquisition procedures were used to increase the probability of a GSR to a given CS. One group of *S*s was conditioned according to the classical Pavlovian paradigm with a CS-UCS interval of .5 sec. A second group was conditioned with a 6-sec. CS-UCS interval. This interval is considerably greater than the established intervals for classical autonomic conditioning, yet it is the interval used by some investigators to establish a heart rate CR (Notterman et al., 1952; Zeaman & Wegner, 1954). A third group was a pseudoconditioning or sensitization group. For *S*s in this group, the CS and UCS were not paired; the UCS was presented randomly during the intertrial interval, never following the CS by less than 30 sec. nor preceding the CS by less than 10 sec.

METHOD

Subjects.—The *S*s were 60 male undergraduate students from the University College of Arts and Science and the College of Engineering of New York University.

Apparatus.—The CS was a 750-cps, 2-sec. tone delivered through earphones at approximately 20 db. The tones were on tape and the presentation of the tones served to program the intertrial intervals. A sound relay and two Hunter timers were used to control the interstimulus intervals and the duration of the CS and UCS. The UCS was a 6-sec. tetanizing shock to *S*'s left arm biceps.

The apparatus for measuring GSR was a modification of the unit described by Schlosberg (1954). It consisted of a 50-microampere meter in series with two quarter-sized silver-silver chloride-saline paste electrodes. The circuit was run by a single 1.5-v. flashlight cell which, with the aid of a potential divider and a calibrated resistor, permitted absolute conductance and GSR in terms of change in conductance to be read directly. In order to insure consistent electrode placement for each *S* and to maintain optimal contact during the experimental session, the electrodes were seated in rubber suction cups which in turn were seated in 4-in.-square wooden blocks. The whole unit was attached to *S*'s palms by means of ½-in. elastics.

The *S*s were seated in a totally darkened booth; an exhaust fan maintained air circulation and served as a masking noise. The *E* and the apparatus were in an adjoining booth.

Procedure.—The *S*s were assigned randomly to the experimental conditions. All *S*s, within each condition, were given an adaptation series consisting of 15 presentations of the CS. Just before beginning the adaptation series *S* was given the following instructions:

"This is an experiment to determine your reactions to tone and to electric shock. You will hear a series of tones through this headset which will soon be placed over your ears. A little later in the experiment, you will receive a series of electric shocks administered to your left arm; you will not be shocked through the electrodes I am now placing on your palms. The shocks you will receive will be strong enough for you to feel them, but will not be strong enough to do you any harm. Remember that they will just pass through your arm and not your body. Please try not to make any unnecessary movements. The apparatus is necessarily very sensitive, and your moving around too much would interfere with the recording."

The adaptation series was followed by the acquisition series consisting of 10 CS-UCS pairings. The interval between the CS and UCS depended upon the condition to which *S* was assigned. Three interstimulus intervals were used: a .5-sec. interval, a 6-sec. interval, and a random interval. In the latter condition the CS and UCS were never paired in that the CS was presented randomly at some point within the interval beginning 30 sec. after the UCS and ending 10 sec. before the next CS.

Following the acquisition series, half of the *S*s in each interstimulus condition were given the following instructions: "From now on you will not be shocked, so simply relax." These instructions were given by *E* who entered *S*'s booth. The remaining 10 *S*s in each of the conditions were given no instructions. All *S*s were given 15 extinction trials. The intertrial intervals for the adaptation, acquisition, and extinction series were programmed on tape and were the same for all *S*s. The intervals ranged from 50 to 75 sec.

RESULTS AND DISCUSSION

The results were tabulated in terms of the mean GSR per *S* for each condition. These means were transformed into logarithms to reduce the marked heterogeneity of variance found in the raw data, and then means for each group were obtained. Table 1 shows the means and *SD*s of the transformed data for adaptation, acquisition, and extinction. An analysis of variance of the adaptation data indicated no significant differences among the groups, thus justifying the assumption that there were no systematic differences in response

TABLE 1

Means and *SD*s of the Transformed GSRs for Adaptation, Acquisition, and Extinction

Extinction Conditions	Meas.	Adaptation			Acquisition			Extinction		
		Training Conditions			Training Conditions			Training Conditions		
		.5 Sec.	6 Sec.	C	.5 Sec.	6 Sec.	C	.5 Sec.	6 Sec.	C
Noninstructed	Mean	1.65	1.79	1.74	2.29	2.27	1.58	2.49	1.88	1.84
	SD	.18	.17	.20	.07	.10	.28	.09	.21	.19
Instructed	Mean	1.80	1.83	1.63	2.21	2.42	1.68	1.63	2.04	.94
	SD	.17	.14	.25	.13	.08	.28	.17	.14	.22

to the CS prior to the experimental treatments.

Table 2 presents the analysis of variance for the acquisition series. There were no significant differences during acquisition between the groups assigned to the different instruction conditions, since the instructions were not given until the extinction series. The differences among the three CS-UCS intervals during acquisition were significant. These differences were between the .5-sec. interval and the control, and between the 6-sec. interval and the control. The .5-sec. and the 6-sec. groups did not differ. Paired *t* tests indicated that the acquisition GSRs of both the .5- and 6-sec. groups were significantly different from the adaptation GSRs (.01 level).

An examination of Table 1 reveals that the increase in GSR amplitude in the 6-sec. group was short-lived in contrast to the increased GSRs found in the .5-sec. group. The mean of the uninstructed 6-sec. group falls back to the adaptation level during extinction, whereas the mean of the .5-sec. uninstructed group remains significantly higher than the adaptation mean (.01 level). This suggests that the .5-sec. group is conditioned, while the 6-sec. group is either weakly conditioned or merely sensitized to the impending shocks. The pseudoconditioning control group shows no reliable change in the mean GSR as a function of the acquisition procedures.

TABLE 2

Analysis of Variance of the Acquisition Data

Source	*df*	*MS*	*F*
Instructions	1	576.60	.16
CS-UCS Interval	2	30,379.65	8.43*
Interval × Instructions	2	673.55	.19
Within cells (error)	54	3,603.84	
Total	59		

* Significant at .01 level.

In order to control for the effects of the acquisition variables in analyzing the extinction data, the extinction means were adjusted in terms of the acquisition means and an analysis of covariance was performed. The analysis of covariance for the extinction series is presented in Table 3, where it can be seen that the instructions significantly affected the GSRs. The interaction calls attention to the fact that the instructions effect was not the same for each of the acquisition CS-UCS intervals. The .5-sec. group was markedly affected by the instructions, as was the control group. In both cases the GSRs were significantly reduced in amplitude (.01 level). However, the

TABLE 3

Analysis of Covariance of the Extinction Data

Source	*df*	*MS*	*F*
Instructions	1	43,333.39	12.90*
CS-UCS Interval	2	23,667.11	7.05*
Interval × Instructions	2	17,477.93	5.20*
Within cells (error)	53	3,358.99	
Total	58		

* Significant at the .01 level.

6-sec. group did not appear to be so affected by the instructions. This latter finding is difficult to explain, in view of the fact that the control group responded to the instructions with diminished GSRs.

One possible interpretation is that for the .5-sec. and control groups, the instructions that no further shocks will be given tends to cancel the initial anxiety-arousing instructions given to all *S*s at the outset of the experiment. If the initial instructions elicit an increase in GSR amplitude, as Mowrer's findings (1938) suggest, then overriding the initial instructions with later information should lower the GSR response. However, it is possible that the 6-sec. interval acquisition procedures produce different behavior, in that *S*s develop a heightened state of GSR response which is no longer dependent upon the initial instructions. If this is the case, then the extinction instructions would not affect the GSR amplitude. In other words, the 6-sec. CS-UCS interval may act as a potent anxiety-arousing technique.

The fact that the 6-sec. interstimulus group did not evidence strong conditioning differs from the findings of Notterman et al. (1952), Zeaman et al. (1954), and Rodnick (1937), but these results are in agreement with those of White and Schlosberg (1952), as well as those of Moeller (1954). In both the Notterman and the Zeaman studies the CR was a measured change in heart rate during a 6-sec. interstimulus interval. Notterman et al. (1952) report a deceleration in rate as the CR, while the data of Zeaman et al. (1954) suggest that the form of the CR depends upon the point at which heart rate is measured. In both cases the possibility exists that the heart rate CR is an indirect effect of the preparatory skeletal responses occurring during the 6-sec. interval.

SUMMARY

Three groups of 20 *S*s were conditioned to give a GSR to a 750-cps tone under different CS-UCS intervals. The intervals were: (*a*) .5 sec., (*b*) 6 sec., and (*c*) a random control interval in which CS and UCS were not paired. Half of the *S*s in each group were given instructions prior to extinction that the UCS (shock) was not going to be given. The results indicated that the instructions affected both the .5-sec. group and the pseudoconditioning control group. The 6-sec. group was not affected by the instructions.

REFERENCES

Cook, S W., & Harris, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21,** 202–210.

Moeller, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48,** 162–166.

Mowrer, O. H. Preparatory set (expectancy)—a determinant in motivation and learning. *Psychol. Rev.*, 1938, **45,** 62–91.

Notterman, J. M., Schoenfeld, W. N., & Bersh, P. J. A comparison of three extinction procedures following heart rate conditioning. *J. abnorm. soc. Psychol.*, 1952, **47,** 674–677.

Rodnick, E. H. Characteristics of delayed and trace conditioned responses. *J. exp. Psychol.*, 1937, **20,** 409–425.

Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61,** 81–88.

White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43,** 357–362.

Zeaman, D., & Wegner, N. The role of drive reduction in classical conditioning of an autonomically mediated response. *J. exp. Psychol.*, 1954, **48,** 349–354.

(Received March 16, 1959)

Journal of Experimental Psychology
Vol. 59, No. 2, 1960

PARTIAL AND CORRELATED REWARD IN ESCAPE LEARNING[1]

GORDON H. BOWER[2]

Yale University

Previous studies (Campbell & Kraeling, 1953; Bower, Fowler, & Trapold, 1959) have demonstrated that magnitude of reinforcement operates as a variable in instrumental escape learning. In those experiments rats were run in a straight alley in which the startbox and runway provided a continuous and intense electric shock, and the goalbox provided a continuous but weaker shock. It was found that running speed was faster the greater the reduction in shock intensity at the goal. Moreover, these performance effects were reversible; when the amount of shock reduction was changed, performance shifted to the level appropriate to the new amount of shock reduction.

Such studies have involved conditions in which *S* receives the same reward on every trial; however, of equal interest are those conditions in which the reward varies from trial to trial. Thus, a large reward (i.e., large amount of shock reduction) may be given on some trials and a small reward on others; or reward may be given immediately on some trials and delayed on others. Experiment I is an investigation of partial reinforcement which may be considered as a limiting case of variable magnitude of reward. Of further interest are those conditions in which variations in the reward are contingent upon variations in some characteristic of the response (e.g., its speed). In Exp. II, one condition of correlated reinforcement is investigated.

EXPERIMENT I

In this experiment the reward was varied independently of *S*'s performance. Two goal events were used, 100% or 0% reduction in shock intensity. Subjects received random sequences of these goal events, with the sequences differing in the proportion of 100% shock reduction trials. These conditions are somewhat analogous to partial reinforcement using, say, food and no food as the goal events.[3]

The primary interest in this experiment was to ascertain the relationship between speed and percentage of reinforcement when the goal events were given in a random sequence. Four groups of *S*s were trained with sequences corresponding to 100, 75, 50, or 25% reinforcement. A second question of interest was whether partial reinforcement in this escape situation causes slow adjustment of performance when the conditions are changed to 0% rewards, which is analogous to extinction.[3]

---

[1] This experiment was conducted during the writer's tenure on a predoctoral fellowship, MF-8585, granted by the National Institutes of Mental Health, Bethesda, Maryland.

[2] Now at Stanford University.

[3] Since a trial in the alley required eventual removal of *S* from the goalbox, the 0% shock reduction event may be described as a long delay of reward (removal from charged goalbox) instead of "nonreward," and as such might maintain performance above operant level. This description in terms of variable delay in no way changes the logic of the experiment.

Method

Subjects.—The *S*s were 20 male albino rats, about 100 days old. They were assigned randomly to four groups of five *S*s and were maintained with free access to water and Purina Lab Checkers.

Apparatus.—The apparatus is described in detail elsewhere (Bower et al., 1959). A 4-ft. straight alley with start and goal compartments was used. Two L-shaped strips of galvanized sheet metal, separated by a 1-in. floor gap, formed the two walls and halves of the floor. Two independent shock systems were connected to the sides of the alley; one shock system provided shock to the startbox and middle section of the alley, the other to the goalbox. The *S* received shock as long as its feet touched the two strips of sheet metal forming the halves of the floor. The 3 in. width of the alley prevented *S* from escaping shock by running along only one side of the alley.

Raising the start door operated a microswitch which closed the shock circuits to the alley and started a Standard Electric clock. The clock stopped when *S* interrupted a photobeam 2 in. inside the goalbox. Running times were recorded to the nearest .01 sec. and were converted to speed scores (1/time).

Design and procedure.—The shock intensity in the startbox and runway was always 250 v. The shock in the goalbox was either 250 or 0 v. Twenty acquisition trials were given. The four groups differed in the proportion (.25, .50, .75, or 1.00) of trials on which the goalbox shock was 0 v.; within these restrictions, random sequences of goal shocks were used. After the initial 20 trials, two groups (100% and 50%) were given 20 additional trials with the 250-v. goal shock present on every trial.

The *S* was placed in the start box and when it oriented toward the goal the start door was raised, turning on shock. When *S* entered the goalbox and interrupted the photobeam, the goalbox door was lowered and *S* was left in the goalbox for 20 sec. The trials were massed; immediately following removal from the goalbox, *S* was reinserted into the startbox for the next trial.

Results

The results in Fig. 1 show group mean speeds plotted in running averages of four training trials. Speed on the later acquisition trials is clearly related to percentage of reinforcement. Analyses of variance of mean speeds on Trials 13–20 demonstrate significant differences between the groups ($F = 12.16$, $df = 3,16$, $P < .001$). Component analysis of the group differences reveal that speed was a linear function of percentage of reinforcement (linear $F = 36.05$, $df = 1,16$, $P < .001$; residual $F < 1.00$). This linear relationship is shown graphically in Fig. 2, where a 0% control group from a previous experiment is also presented.

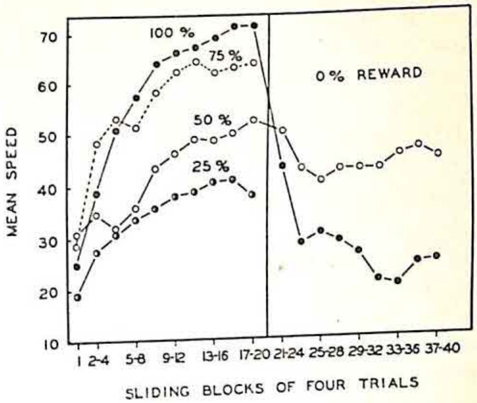


FIG. 1. Group mean speeds in sliding blocks of four trials. After Trial 20, the 100% and 50% groups were switched to a 0% reward condition.

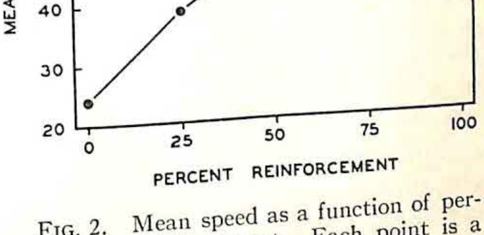


FIG. 2. Mean speed as a function of percentage or reinforcement. Each point is a group average speed over Trials 13–20.

The results of shifting to 250-v. goal shocks on every trial may be seen in the right half of Fig. 1. The *S*s trained under 50% reinforcement show considerable resistance to change in performance, whereas *S*s trained under 100% reinforcement show an abrupt decrement in speed when switched to 0% reward. The speed curves for the two conditions cross after only four trials at the new shock condition. A trend analysis over Trials 17–40 shows a reliable groups-by-trialsinteraction ($F = 7.05$, $df = 5{,}40$, $P < .001$), which demonstrates the difference in rate of shift.

EXPERIMENT II

In Exp. II, variations in reward were correlated with running speed, reward being given when *S* responded slowly. In the present study, the cutoff selected was a speed of 40, or running time of 2.5 sec. and 100% shock reduction was given for speeds below 40 and 0% shock reduction for speeds above 40. This condition is analogous to the DRL schedule studied extensively in the Skinner box (Ferster & Skinner, 1957).

Logan (1959) has proposed an equilibrium model for conditions of correlated reinforcement. The equilibrium performance is defined as that speed which, on the average, gets that percentage of reinforcement which will maintain just that average speed. The most direct test of this model is to compare the performance of the correlated reward *S*s with that of matched control mates. The matched control *S* receives the same sequence of rewards as that received by its experimental partner; however, these rewards are given irrespective of the control *S*'s speed. According to the equilibrium model, performance depends only upon the sequence of rewards and not upon the correlation of reward with speed. Because the sequence of rewards is equated, the equilibrium prediction is that correlated *S*s and their matched controls will perform similarly.

Method

Four pairs of rats were used, with the correlated and matched control *S*s of the pairs designated randomly. The apparatus and general procedure was the same as in Exp. I. The hair was shaved from *S*s' backs to provide better electrical contact. Each *S* received 50 massed trials. For an *S* in the correlated group, the 250-v. goal shock was turned off as soon as it broke the photobeam provided its running time exceeded 2.50 sec.; otherwise, the goal shock was left on. In either case, *S* was removed from the goalbox after 20 sec. The matched control procedure insured that a correlated *S* and its control partner were equated on the entire sequence of goal events.

Results

Group mean speeds in blocks of five trials are shown in Fig. 3. The curves start and rise together for 10 trials, but thereafter show progressive separation. The separation on Trials 41–50 was statistically significant ($F = 55.25$, $df = 1{,}6$, $P < .001$). The mean speed for correlated *S*s rises initially above the cutoff speed

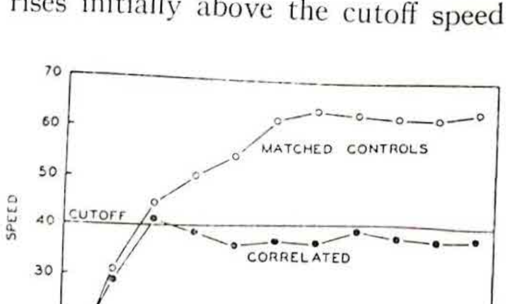


FIG. 3. Group mean speeds in blocks of five trials. The horizontal line represents the cutoff speed of 40.

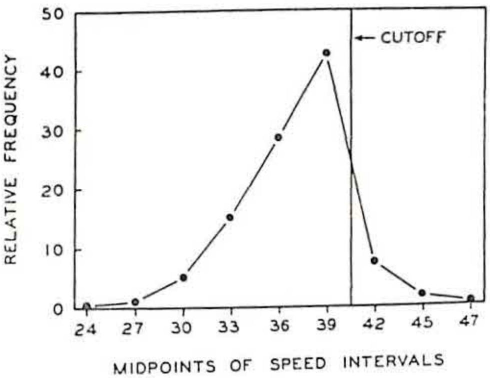


FIG. 4. Pooled speed distribution for the correlated group; plotted by intervals of three speed points.

of 40, but then declines to a relatively stable average speed just below the cutoff. The mean speeds over Trials 41–50 of the four *S*s in the correlated group were 37, 36, 36, and 38; their respective matched controls had mean speeds of 71, 65, 56 and 58. Apparently, the correlated condition reduced between-*S* variability to a large extent.

The pooled distribution of speeds over Trials 26–50 for correlated *S*s is shown in Fig. 4. A distinct mode appears in that speed interval just below the cutoff. Over the last 25 trials about 90% of their runs were rewarded.

DISCUSSION

The results of Exp. I show that speed is a linear function of the percentage of reinforcement; that is, the speed produced by random variations between the two rewards (100% and 0% shock reduction) is approximately the average of the speeds produced by the two rewards separately. Although this result was obtained with escape learning, it is consistent with the results of Logan[4] for variable magnitude and delay of food reinforcement. The result is contrary to the results obtained by Weinstock (1958) and Goodrich (1959) by varying percentage of food reward and nonreward. However, a number of procedural differences (e.g., number and massing of trials, nonrewards vs. long delays, the part of the response chain being measured, etc.) make serious comparisons difficult. That performance changes are slower following training with variable reward in this situation is consistent with the general findings for instrumental appetitive conditioning

The results of Exp. II clearly disconfirm the equilibrium prediction. Performance depends not only upon percentage of reinforcement but also upon the correlation of reward with speed. It should be noted that the performance of the uncorrelated, matched control *S*s is close to the speed expected on the basis of previous findings. Over-all, the percentage of reinforcement for the matched controls was about 80–85%. From Fig. 2, this percentage of reinforcement, when given uncorrelated with speed, would be expected to produce an average speed of about 64. This figure compares favorably with the average of 63 observed for the matched control group.

The fact that correlated *S*s learned to run slowly is consistent with the micromolar theory proposed by Logan (1956; 1959). In this approach, different speeds are treated as different responses, and the probability of a particular speed depends upon its reward relative to the rewards for competing speeds. The reinforcement dimension most relevant to the present results is the interval of reinforcement, defined as the time between the initiation of the response and the receipt of reward. The incentive for a particular response speed is assumed to be a decreasing function of its interval of reinforcement. For the present contingencies, this principle clearly favors speeds below 40 since they receive the reward immediately. Furthermore, although speeds of 38 and 20, for example, receive the same goal events, speeds of 38 have higher incentive because they have a shorter interval of reinforcement. In other words, speeds of 20 are associated with 5 sec. of shock in the alley, whereas speeds of 38 are associated with only 2.65 sec. of alley shock. Thus, it is expected that *S* will learn to make predominately those response chains which minimize shock time, and the predicted mode of the speed distribution in Fig. 4 is at that speed interval just below the cutoff. Because the controlling contingencies "drive" each *S* into this optimal response region, the between-*S* differences are negligible.

---

[4] Results reported in a forthcoming book, to be titled *Incentive*.

SUMMARY

Rats were trained to run down a straight alley to escape a 250-v. shock, the shock in the goalbox being either 0 or 250 v. In Exp. I the percentage of reinforcement was either 25, 50, 75, or 100, and it was found that speed was a linearly increasing function of this variable. When the 50% and 100% groups were switched to 0% reward, the 100% group extinguished faster than the 50% group.

In Exp. II reward was correlated with slow speeds, the goal shock being 0 v. for speeds below 40 and 250 v. otherwise. Matched control *S*s received the same sequence of goal events as that received by correlated *S*s, but reward was uncorrelated with the control *S*'s speed. It was found that *S*s in the correlated group learned to run at speeds just below the cutoff, receiving about 90% reward, whereas the matched controls ran significantly faster. This latter result was related to a micromolar theory which considers interval of reinforcement as one component of incentive motivation.

REFERENCES

BOWER, G. H., FOWLER, H., & TRAPOLD, M. A. Escape learning as a function of amount of shock reduction. *J. exp. Psychol.*, 1959, **58**, 482–484.

CAMPBELL, B. A., & KRAELING, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 97–101.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–64.

LOGAN, F. A. A micromolar approach to behavior theory. *Psychol. Rev.*, 1956, **63**, 63–73.

LOGAN, F. A. The Hull-Spence approach. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. II. New York: McGraw-Hill, 1959.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151–158.

(Received March 20, 1959)

Journal of Experimental Psychology
Vol. 59, No. 2, 1960

THE EFFECT OF STIMULUS SIMILARITY ON AMOUNT OF CUE–POSITION PATTERNING IN DISCRIMINATION PROBLEMS

BARBARA NOTKIN WHITE[1] AND CHARLES C. SPIKER

Iowa Child Welfare Research Station

In the psychological literature, much attention has been paid to the problem of defining the effective stimuli in discrimination problems. Some of the discussion has been concerned with whether the response is conditioned to each of the individual components of a compound confronting *S*, or is conditioned to the compound per se. The early theory of discrimination behavior proposed by Spence (1936) and the later analysis of discrimination learning by Hull (1952) assumed that the response is conditioned to each component of a compound. Although Spence's (1952) theory has proved to be highly fruitful in explaining a number of phenomena discovered in the study of simultaneous discrimination learning, subsequent extensive study of the so-called successive discrimination problem led him to publish a theoretical analysis of this type of problem. He noted that his earlier theory was not applicable to the successive problem, since no component in the cue-position compounds of the successive problem is consistently reinforced more frequently than any other. He proposed the notion of cue-position patterning to account for the learning that occurs in the successive problem. In effect, he suggested that the response is conditioned to the compound consisting of a cue (e.g., black or white) and a position (e.g., left or right). Thus, *S* learns to discriminate between positive and negative compounds rather than between positive and negative components. Furthermore, Spence suggested that cue-position patterning occurs only when there is nondifferential reinforcement of the components (i.e., no one of the cue members of the compound is systematically reinforced more than the others).

The notion that response to patterns does occur under conditions where a component is differentially reinforced was directly investigated by Birch and Vandenberg (1952). They analyzed the cue-position compound (e.g., a black left alley) into three aspects—the spatial cue (S_L), the nonspatial cue (S_B), and a patterned stimulus (the compound) (S_{BL}). They noted that, in the simultaneous problem, both the nonspatial cues and the patterned cues are differentially reinforced; in the successive problem, on the other hand, only the patterned cues are differentially reinforced. Thus, only the patterned stimuli can serve as the basis of solution of the successive problem, whereas both types of cues might be used in the simultaneous problem. Accordingly, it is meaningful to ask, in connection with the simultaneous problem, to what degree the discrimination is determined by each of the two types of cues.

Birch and Vandenberg devised a convenient way of determining to what degree cue-position patterning develops in a simultaneous problem. Table 1 presents a schematic representation of a procedure suggested by them. In Problem 1, two simultaneous discrimination problems are used; however, only one spatial arrangement of each is presented, with the positive stimulus of the second setting in the position opposite from that of the positive stimulus in the first setting. In Problem 2, the test problem, the spatial arrangements of the two settings are reversed. In Problem 1, it will be noted that the left and right positions are equally

---

[1] Now at Michael Reese Hospital, Chicago.

TABLE 1

REPRESENTATION OF PROCEDURE FOR ISOLATING DEGREE OF CONDITIONING TO STIMULUS COMPONENTS AND TO PATTERN

Problem 1 (Training)	Setting 1	S_LS_A	(+)	S_RS_B	(−)
	Setting 2	S_LS_C	(−)	S_RS_D	(+)
Problem 2 (Testing)	Setting 1	S_LS_B	( )	S_RS_A	( )
	Setting 2	S_LS_D	( )	S_RS_C	( )

Note.—S_L and S_R refer to the left and right spatial positions; S_A and S_B refer to one pair of nonspatial cues (e.g., black and white); S_C and S_D refer to a second pair of nonspatial cues (e.g., red and green); (+) and (−) designate the reinforced and nonreinforced complexes of Problem 1; and ( ) refers to complexes which were neither reinforced nor nonreinforced (see text).

often reinforced and nonreinforced, whereas S_A and S_D are consistently reinforced and S_B and S_C are consistently nonreinforced.

If, in Problem 1, the response is differentially conditioned only to the nonspatial cue components, it should occur without disruption to the nonspatial cues of Problem 2. Thus, in the schema of Table 1, 100% response to S_A and S_D should be expected following mastery of Problem 1. It may be noted that the patterned stimuli of Problem 1 (S_{LA}, S_{RB}, S_{LC}, and S_{RD}) do not occur in Problem 2. Therefore, if the tendency for *S* to respond in Problem 1 has been determined entirely by differential conditioning to the patterned stimuli, he should respond equally often (i.e., by "chance") to the four stimulus compounds of Problem 2. Finally, if the response was differentially conditioned both to the nonspatial components and to the patterned stimuli of Problem 1, there should be a reduction in the tendency to choose S_A and S_D in Problem 2 due to the change in the patterned stimuli; but the choice of these stimuli should exceed 50% owing to the influence of the habits differentially established to the nonspatial cues. From these considerations, it is apparent that the amount of cue-position patterning that occurs in Problem 1 may be measured in Problem 2.

The results of the experiment by Birch and Vandenberg (1952) demonstrate rather convincingly that with rats cue-position patterning occurs in the simultaneous situation, despite the fact that the components of the cue-position compounds receive differential reinforcement.

One of the implications of the Birch and Vandenberg analysis is that the relative influence of the patterned stimuli in such a problem is determined by the relative ease with which differentiation can be achieved by means of the nonspatial components. Thus, in the simultaneous problem, involving highly distinctive nonspatial components within each setting, the role of the patterned stimuli should be negligible. At the other extreme is the successive problem in which the nonspatial components within each setting are identical. In the latter case, discrimination theoretically must be based entirely on differential habits established to the patterned stimuli. This suggests that the amount of cue-position patterning should be an increasing function of the degree of similarity between the nonspatial cues within each setting.

The purpose of the present experiment is to determine whether or not preschool children manifest cue-position patterning in a problem similar to that of Birch and Vandenberg, and to test the additional hypothesis that the amount of cue-position patterning is an increasing function of the similarity of the nonspatial cues within a given trial setting (i.e., the within-setting similarity). While the latter hypothesis is, in effect, a modification of Spence's theory, its confirmation within this procedure can at best be only suggestive, since the theory is explicitly restricted to inarticulate organisms.

METHOD

Experimental design.—In order to test these hypotheses, three groups of *S*s were run, each with a different degree of similarity between the two nonspatial cues presented on each trial (Table 2). For Group I, the stimuli were dissimilar; for Group II, similar; and for Group III, identical. The test situations were the same as the corresponding training situations for each group, except that the spatial position of each cue stimulus was reversed. For each group, differential responding to the components was reflected in *S*'s tendency to select, on each test trial, the previously reinforced nonspatial cue. Conversely, response to patterns was indicated by *S*'s tendency to select, within each setting, the stimulus in the position previously occupied by the positive stimulus of the training problem.

On the assumption that all discrimination problems involve some patterning, and that the relative effectiveness of the patterned stimuli is an increasing function of within-setting stimulus similarity, it was expected that the choice of the stimulus in the previously reinforced position would be very low for the group trained on dissimilar stimuli (Group I), very high for the group in which the stimuli were identical (Group III), and somewhere in between for the group with quite similar stimuli (Group II).

Apparatus.—The stimuli were eight 3 × 3-in. pieces of colored paper backed by heavy white plastic. On every trial, the two stimuli to be used were fitted into the front faces of two 3½-in. cubical wooden boxes mounted 6 in. apart on a 17-in. tray. These were presented to the child on top of a stand 12 in. high. The boxes were separately hinged to the tray on the side away from the stimulus slot so that they could be lifted by *S* to obtain (on reinforced trials) a marble located on a groove in the tray. Each marble so won was placed in a small plastic container. The entire apparatus, with the exception of the marble container, was painted flat black.

Subjects.—The *S*s were 42 children from the Iowa Child Welfare Research Station preschool laboratories, their ages ranging from 4 to 5½ yr. The data of an additional 14 *S*s were discarded; one because of incorrect instructions, the other 13 because of failure to learn the problem in a reasonable number of trials. Of these, five were from Group I, four from Group II, and five from Group III. The *S*s were assigned by a random procedure to each of the three treatment groups; each treatment group finally containing 14 *S*s.

Procedure.—The design of the experiment is presented in Table 2. It can be seen that the only essential difference among the training problems of the three groups lay in the degree of similarity of the stimuli presented on a trial. In the case of the stimuli for Group II, the discriminability between the red and orange-red and between the white and light gray were subjectively judged by several independent observers to be about equivalent. It should also be noticed that the correct stimuli and the sides on which they were placed were the same for all groups.

All *S*s within any group were trained to a criterion of four consecutive correct responses on the type of trial involving red as the reinforced stimulus, and then received training to the same criterion on the second type of trial. Once both criteria were met, further training continued as follows: the different types of trials were alternated according to a predetermined random order which was repeated three times in 36 trials. When necessary, one or two blocks of six trials were interspersed to break up position or alternating habits that appeared to be forming. When *S* reached a subcriterion of 5/6 correct responses in a block of six trials, he was presented with the criterion block of six trials, the order of which differed from the order used in training. If he did not meet

TABLE 2

SCHEMATIC REPRESENTATION OF THE EXPERIMENT

Group	Setting	Train (Problem 1)				Test (Problem 2)			
I	1	S_LS_{red}	(+)	S_RS_{yellow}	(−)	S_LS_{yellow}	( )	S_RS_{red}	( )
	2	S_LS_{black}	(−)	S_RS_{white}	(+)	S_LS_{white}	( )	S_RS_{black}	( )
II	1	S_LS_{red}	(+)	S_RS_{orange}	(−)	S_LS_{orange}	( )	S_RS_{red}	( )
	2	S_LS_{gray}	(−)	S_RS_{white}	(+)	S_LS_{white}	( )	S_RS_{gray}	( )
III	1	S_LS_{red}	(+)	S_RS_{red}	(−)	S_LS_{red}	( )	S_RS_{red}	( )
	2	S_LS_{white}	(−)	S_RS_{white}	(+)	S_LS_{white}	( )	S_RS_{white}	( )

Note.—S_L and S_R refer to the left and right spatial positions; (+) and (−) designate the reinforced and nonreinforced complexes of the training problem; and ( ) refers to complexes which were neither reinforced nor nonreinforced. The stimuli appeared in the spatial positions indicated above.

this criterion of all six correct, training was continued, and a second opportunity on the criterion block allowed. The final performance criterion was the same for all groups.

For each of the three groups, the test situation consisted of 10 presentations of the same stimuli used during training, with one difference—the side on which a given stimulus was placed was changed (see Table 2). The order of presentation of the different types of trials was the same as that used for the latter part of training. However, one-half of the *S*s in each group started the series with one type of trial (Subgroup A), the second half with the other type (Subgroup B); then each followed through the same predetermined random sequence of changes from one type of trial to the next.

Before entering the experimental room, each *S* was allowed to choose one of a selection of dime store toys to serve as a prize. He was told that he would be permitted to play a game in which he could win enough marbles to buy the toy. On each trial, if *S* selected correctly, he found a marble under the box. The noncorrection procedure was used throughout.

During the test trials, no reinforcement was given, so that the tendencies to respond remained relatively unchanged after the training series. In an attempt to avoid the extinctive effects of nonreinforcement, the following method was used: *E* informed *S* that she would remove the marble for him, that she would keep track of the number of marbles he won, and that when the game was over, she would show him how many marbles he had won. During this phase, *S* was asked only to point to the box which he wished to select, and *E* did not inform him whether or not his choice was correct.

RESULTS

Examination of the training data (Table 3) reveals little difference among the three groups in the mean number of trials to criterion. A simple analysis of variance comparing the groups gave a nonsignificant F ($F < 1.0$). Greater difficulty of the successive problem would not necessarily be expected in this paradigm as the successive problem involved only two nonspatial cues while the other two problems each involved four.

TABLE 3

TRIALS TO CRITERION IN TRAINING PROBLEM AND RESPONSES TO PREVIOUSLY REINFORCED POSITION

Group	Trials to Criterion		Patterned Responses	
	Mean	*SD*	Mean	*SD*
I	26.71	14.52	2.36	2.50
II	26.29	11.95	6.50	3.64
III	24.57	17.70	9.36	1.59

Considering the test trials, the responses to the previously reinforced position were tabulated for each treatment group, keeping separate the data of the A and B subgroups (which counterbalanced for the type of test trial first received). Since *t* tests showed no differences between the A and B subgroups within any treatment, these were collapsed in the final analysis of the data. Table 3 gives the means and *SD*s for each treatment group over all 10 test trials. As was expected, the mean number of responses to the previously reinforced side increased as the similarity of the within-pair stimuli increased (i.e., from Group I to Group III). A test suggested by Cochran and Cox (Snedecor, 1948) was used to analyze the results, since the within-groups variance, owing to the almost complete uniformity in the performance of *S*s in Group III, differed among treatment groups. An analysis over all 10 test trials indicated that Groups I and II differed at the .01 level, and that the difference between Groups II and III was also significant ($.05 > P > .02$).[2]

In Groups I, II, and III, the numbers of *S*s responding to position on the first test trial were 3, 9, and 14 respectively. A chi square test comparing these scores gave a value of 18.38, significant at the .001 level for 2 *df*.

---

[2] A Mann-Whitney *U* test was also performed on the data, yielding essentially the same results.

The analyses of all test trials and of the first test trial, considered together, seem to indicate that the effects of training appeared early in the test series and remained relatively constant throughout all 10 test trials.

DISCUSSION

The results of the present study support the contention that the effectiveness of the patterned stimuli in any trial setting (i.e., the degree of patterning) increases as the similarity of the stimuli for that setting increases. In Groups I and II, if the pattern were not an effective cue, *S*s would have continued to respond consistently to the previously reinforced nonspatial stimulus when these stimuli were shifted in the test problem. In the case of Group III, unless response to patterns had occurred, Problem 1 would have been insoluble; since the test problem was the same as the training problem, patterning was again revealed by selection of the originally reinforced pattern. Hence, in each group, response to patterns was manifested in each of the two stimulus settings of the test problem by selection of the stimulus presented in the same position as that of the previously reinforced stimulus. This manifestation was greatest in Group III, and least in Group I.

The data for Group III may be used to determine whether or not the introduction of the novel suspension of reinforcement in the test situation resulted in a general disruption of performance. Since Problem 1 was the same as Problem 2 for this group, any marked change in performance between the two problems could be interpreted as the result of changes in the conditions under which *S*s performed. The performance of Group III, however, showed very little change from Problem 1 to Problem 2. This fact, together with the fact that the analysis of the first test trial leads to the same conclusion as does the analysis of all 10 test trials, indicates that the introduction of the test did not disrupt performance.

The possibility remains, and must be dealt with, that some of the results reported in support of the hypothesis are artifacts of the experimental design and the analyses used.

Firstly, it is possible that the increasing amounts of cue-position patterning with increasing within-setting stimulus similarity reflected not individual behavior, but rather group tendencies, and were the results of averaging over *S*s, some of whom showed response to nonspatial cues only, some of whom responded to patterns only. If this were so, then increase in patterning with increase in similarity would be due to an increase in the proportion of *S*s who showed all (as opposed to no) patterning. That such a factor does not entirely account for the results is suggested by the following tabulations: in Group I, only two *S*s gave as many as 6 responses to patterns, while nine gave 0 or 1. In Group III, only two *S*s showed less than 10 such responses, and neither of these two showed less than 5; in the case of Group II, three *S*s showed between 0 and 2 responses to patterns, five gave between 3 and 7, and six *S*s gave between 8 and 10.

Another possibility directly concerns Group II where an intermediate number of responses to patterns was obtained. To rule out the possibility of this having been obtained by the majority of *S*s responding consistently to one position on all 10 trials, two analyses were made. The number of *S*s giving between 8 and 10 responses to one side was three, while the majority (11) showed between 5 and 7 responses to any one side (hence, between 5 and 3 to the other). This corresponds with Groups I and III where the numbers of *S*s (out of 14 in each case) responding 8 or more times to one position were 4 and 1 respectively. Also, if *S*s responded consistently to one position in all 10 test trials, they would of necessity show "response to patterns" consistently to one of the two settings and no patterning-type responses to the other setting. Hence, another way of tackling the same problem is to correlate the numbers of patterned responses in Setting 1 with those in Setting 2. For the 14 *S*s in Group II, the Pearson r was .672 ($P < .01$), indicating that those *S*s who patterned low or high in one setting tended to respond similarly in the second setting. (The same analysis over all 42 *S*s yielded a correlation of .635 ($P < .001$), but this is probably inflated by the preponderance of 5's for both settings in Group III).

Thus it would seem that the results obtained are indicative of increasing responses to cue-position patterns with increasing stimulus similarity, and are not simply an artifact of averaging or of *S*s adopting a position preference on the test trials.

Some of the work done by Bitterman and his associates is relevant to the thesis presented here. Teas and Bitterman (1952) reported data which correspond exactly to the predictions generated by the hypothesis of this paper. In addition, a statement by Bitterman and Wodinsky (1953, p. 375), although phrased in different terminology and couched in a different theoretical framework, appears to be, at least in part, a statement of the hypothesis proposed here.

SUMMARY

This experiment studied the effects of increasing the similarity of the within-pair stimuli in a discrimination problem on the amount of patterned responding in a nonreinforced test series. Three groups of preschool children were trained to respond to a red stimulus on the left side on one type of trial and to a white stimulus on the right side on a second type of trial in a discrimination situation involving the choice of one or the other of two boxes fronted by colored cue stimuli. For one group, the nonreinforced cue stimuli were quite dissimilar to the rewarded stimuli with which they were paired. For another, the dissimilarity between each pair was less, and for the last, the two stimuli presented on a trial were identical. It was found that in the subsequent test series, which involved the training stimuli in reversed spatial position, the number of responses to the side previously correct for a given stimulus setting (i.e., patterned responding) increased significantly as the similarity increased.

REFERENCES

BIRCH, D., & VANDENBERG, V. The necessary conditions for cue-position patterning. *J. exper. Psychol.*, 1955, **50**, 391–396.

BITTERMAN, M. E., & WODINSKY, J. Simultaneous and successive discrimination. *Psychol. Rev.*, 1953, **60**, 371–376.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

SNEDECOR, G. W. *Statistical methods.* Ames: Iowa State Coll. Press, 1948.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

SPENCE, K. W. The nature of response in discrimination learning. *Psychol. Rev.*, 1952, **59**, 89–93.

TEAS, D. C., & BITTERMAN, M. E. Perceptual organization in the rat. *Psychol. Rev.*, 1952, **59**, 130–140.

(Received April 7, 1959)

Journal of
Experimental Psychology

VOL. 59, NO. 3 MARCH 1960

THE ROLE OF MEMORY IN THE ACQUISITION OF CONCEPTS[1]

HUGH E. CAHILL[2] AND CARL I. HOVLAND

Yale University

In the learning of concepts under naturalistic conditions there is typically a time interval between successive instances and also between the presentation of instances and the testing of concept attainment. An important research problem is thus generated concerning the effect of memory in the acquisition of concepts. This problem also is of significance in the simulation of cognitive behavior by high speed computers (Hovland, 1959). With computer simulation the first approximations often involve conditions of unlimited memory. In human learning, however, forgetting, due to interference and other factors, is typical and must be understood and incorporated into the model if programing of computers is to simulate human performance successfully.

In the present experiment the effect of memory factors is analyzed by comparing performance in acquiring concepts under conditions simulating unlimited memory storage with that under conditions more closely simulating human learning where forgetting of previous exposed instances is compounded with the inference phase of concept acquisition. Instances were successively exposed, the "unlimited memory" condition being simulated by allowing all of the instances to remain in view and the "limited memory" condition by allowing only the last presented instance to remain exposed.

METHOD

The concept model utilized by Hovland (1952) and Hovland and Weiss (1953) was employed. This involves a "closed" system in which the character of the concept to be learned is specified to *S*, i.e., he is told what dimensions are involved, how many are relevant, and the number of values for each dimension. The concept instances were constructed with three values within three stimulus dimensions: form, (triangle, circle, or cross), shading (white, gray, or stripe), and number (one, two, or three). In the present study the concept was transmitted entirely by negative instances.

For the practice series the concept involved only one value of one dimension. An example of such a concept would be *circu-*

---

[1] Research funds were generously provided by the Behavior Science Division of the Ford Foundation. The study was conducted during 1953–54 and preliminary results were reported at the APA meetings, September 1954. The authors wish to thank Fred Sheffield for careful reading of the manuscript.

[2] Now at Psychological Research Associates, Inc., Encino, California.

larity. Negative instances would be those containing one, two, or three white, gray, or striped crosses or squares. Three negative instances appropriately chosen would suffice to exclude all but the correct concept.

For the experiment itself two more complex conjunctive concepts were used. One, hereafter designated as a Level 1 concept, involved two relevant dimensions (one irrelevant) with one value out of three correct for each dimension (e.g., a white circle). Twenty-seven concepts of this type may be generated. Each negative instance excludes three concepts but due to redundancy 10 negative instances are required to exclude all hypotheses except the correct one.

The other, more complex two-valued concept used in the main experiment involved two relevant dimensions (one irrelevant) with two values out of three correct for each dimension (e.g., white and gray circles and triangles). This problem will be hereafter described as one of Level 2. Again 27 concepts can be generated. Here, however, each negative instance eliminates 12 possible concepts. Four negative instances eliminate all but the correct concept (again there is some redundancy necessary).

A "guess-the-correct-concept" technique was employed. The *S*s wrote down after each instance was presented their best guess as to the concept involved, thus providing a continuous record of the types of hypotheses used by *S*. A practice booklet and test booklet were devised which contained all of the necessary experimental information given to *S*s. The practice booklet contained three pages of general orientation instructions which indicated by verbal definition and pictorial examples the stimulus universe, the nature of negative concept instances and the structure of the concepts to be employed. To permit *S*s to test in practice their assimilation of these materials two concept problems, one of Level 1 and one of Level 2, were also included in each practice booklet.

Simultaneous ("unlimited memory") series. Here new instances were exposed at 30-sec. intervals, with the instance remaining in view once it had been exposed. Each page containing all previous instances plus the new one. Answer sheets, on which *S* was required to record his tentative concept hypotheses, were alternated with each of the pages portraying instances.

Successive ("limited memory") series. This method also involved exposing each instance for 30 sec., but at the end of this period the instance was no longer available. This was effected by exposing only the new instance on successive pages of the booklet. Answer sheets were again interspersed between pages of instances.

During practice a Level 1 problem was presented by the successive method. A Level 2 problem was given by simultaneous presentation and also by the successive method.

During the main experiment two Level 1 problems and two Level 2 problems were presented. One problem at each level was given by the simultaneous presentation ("unlimited memory") and one by the successive presentation ("limited memory") technique.

The concepts in the practice and in the learning series differed in the particular combination of dimensions and values which they contained. In all, three concepts were devised for each level of problem complexity so that each *S* received one problem of each level during practice and the remaining two in the learning-testing phase. The problems employed were so arranged that each third of the total group of *S*s received each type of practice problem.

Twice the minimal required number of instances was given for each problem in an attempt to insure a reasonably high level of performance while at the same time extending the opportunity to obtain a record of *S*'s intervening tentative concept hypotheses. No indication was given to *S*, however, as to when a sufficient amount of information for deducing the correct concept had been presented. Accordingly it was assumed that *S*s looked upon the instances presented as forming one continuous series.

Subjects. Thirty-six Yale undergraduates hired through the university employment bureau served as *S*s in this experiment. The *S*s were tested in groups of two to five depending on their availability. Two *S*s who were unable to learn the Level 1 problem even under simultaneous conditions were replaced, as were seven *S*s who failed to follow instructions, such as those requiring *S*s to give a guess after each instance.

After each *S* was given a practice and experimental booklet *E* read aloud the three pages of orientation material while *S*s read the same material silently. In this material *S* was exposed to the stimulus universe and was shown a number of instances of concepts involving a single value each (practice-level problems) and was introduced to the role of negative instances as they are used in presenting information pertinent to simple concept problems. The *S*s were then asked to solve three practice-level problems. They typically accomplished this in 2 min. or less, thus indicating their assimilation of the explanatory materials given them to that point. Subsequently *S*s were practiced on problems of Levels 1 and 2. Since all of the relevant experimental materials were contained in booklets it was possible to have each *S* work on a different problem during the same experimental session.

Following the practice session *E* emphasized the importance of the standardized test conditions which required *S*s to turn the pages in their test booklets exactly on cue at the end of each 30-sec. period. Specifically 20 sec. were allowed for perusal of the instances and up to 10 sec. for recording of tentative concept hypotheses. No questions were permitted during the learning-testing phase and *S*s were given their four test problems consecutively with a rest period of about 2 min. intervening between problems.

RESULTS

The extent to which memory plays an important role in concept learning is shown by the comparison of the number of correct statements of the concepts after the presentation of the minimum number of instances (after Trial 10 for Level 1 problems and after Trial 4 for Level 2 problems) under the simultaneous and successive conditions. Eleven test trials were involved in the former series (after Instances 10 through 20) and five in the latter (after Instances 4 through 8). Success scores determined on this basis were not normally distributed and tests of significance of success as a function of type of stimulus presentation and problem complexity level were made by the Wilcoxon (1945) nonparametric test for paired-replicates. The results of these tests, as well as the median number of correct responses, for the different stimulus presentation conditions are summarized in Table 1. Here it may be seen that concept attainment under simultaneous presentation is markedly superior at both levels of problem difficulty ($P < .001$ for Level 1 and $P = .02$ for Level 2). The findings replicate those of Hovland and Weiss (1953) in showing poorer performance when previous instances must be retained in memory rather than being directly available for inspection.

TABLE 1

NUMBER OF CORRECT STATEMENTS OF CONCEPT AS A FUNCTION OF TYPE OF STIMULUS PRESENTATION AND PROBLEM

Condition	No. Nonzero Differences	Median	Diff.	One-Tail P
Level 1 problem				
Simultaneous	34	10.8	10.4	.001
Successive		0.4		
Level 2 problem				
Simultaneous	28	4.6	4.2	.02
Successive		0.4		

Note.—Level 1 problem involved instances 10 through 20, with maximum score of 11; Level 2 problem involved instances 4 through 8, with maximum score of 5.

Errors were analyzed to determine how many were responses incompatible with the immediately presented instance and how many were incompatible with instances presented somewhere earlier in the series. The former may be described as "perceptual-inference" errors since they are responses to stimuli physically available and constitute failure either to perceive correctly the stimulus presented or to make the appropriate inference as to possible concepts from it.

Perceptual-inference errors were reltively infrequent. With the Level 1 problem 5 *S*s made a total of 6 errors under simultaneous conditions and 7 *S*s made a total of 8 errors under successive conditions. With the Level 2 problem 12 *S*s made a total of 24 errors under simultaneous conditions and 15 *S*s made a total of 21 under successive. As would be expected, no reliable difference between simul-

larity. Negative instances would be those containing one, two, or three white, gray, or striped crosses or squares. Three negative instances appropriately chosen would suffice to exclude all but the correct concept.

For the experiment itself two more complex conjunctive concepts were used. One, hereafter designated as a Level 1 concept, involved two relevant dimensions (one irrelevant) with one value out of three correct for each dimension (e.g., a white circle). Twenty-seven concepts of this type may be generated. Each negative instance excludes three concepts but due to redundancy 10 negative instances are required to exclude all hypotheses except the correct one.

The other, more complex two-valued concept used in the main experiment involved two relevant dimensions (one irrelevant) with two values out of three correct for each dimension (e.g., white and gray circles and triangles). This problem will be hereafter described as one of Level 2. Again 27 concepts can be generated. Here, however, each negative instance eliminates 12 possible concepts. Four negative instances eliminate all but the correct concept (again there is some redundancy necessary).

A "guess-the-correct-concept" technique was employed. The *S*s wrote down after each instance was presented their best guess as to the concept involved, thus providing a continuous record of the types of hypotheses used by *S*. A practice booklet and test booklet were devised which contained all of the necessary experimental information given to *S*s. The practice booklet contained three pages of general orientation instructions which indicated by verbal definition and pictorial examples the stimulus universe, the nature of negative concept instances and the structure of the concepts to be employed. To permit *S*s to test in practice their assimilation of these materials two concept problems, one of Level 1 and one of Level 2, were also included in each practice booklet.

Simultaneous ("unlimited memory") series. Here new instances were exposed at 30-sec. intervals, with the instance remaining in view once it had been exposed. Each page containing all previous instances plus the new one. Answer sheets, on which *S* was required to record his tentative concept hypotheses, were alternated with each of the pages portraying instances.

Successive ("limited memory") series. This method also involved exposing each instance for 30 sec., but at the end of this period the instance was no longer available. This was effected by exposing only the new instance on successive pages of the booklet. Answer sheets were again interspersed between pages of instances.

During practice a Level 1 problem was presented by the successive method. A Level 2 problem was given by simultaneous presentation and also by the successive method.

During the main experiment two Level 1 problems and two Level 2 problems were presented. One problem at each level was given by the simultaneous presentation ("unlimited memory") and one by the successive presentation ("limited memory") technique.

The concepts in the practice and in the learning series differed in the particular combination of dimensions and values which they contained. In all, three concepts were devised for each level of problem complexity so that each *S* received one problem of each level during practice and the remaining two in the learning-testing phase. The problems employed were so arranged that each third of the total group of *S*s received each type of practice problem.

Twice the minimal required number of instances was given for each problem in an attempt to insure a reasonably high level of performance while at the same time extending the opportunity to obtain a record of *S*'s intervening tentative concept hypotheses. No indication was given to *S*, however, as to when a sufficient amount of information for deducing the correct concept had been presented. Accordingly it was assumed that *S*s looked upon the instances presented as forming one continuous series.

Subjects. Thirty-six Yale undergraduates hired through the university employment bureau served as *S*s in this experiment. The *S*s were tested in groups of two to five depending on their availability. Two *S*s who were unable to learn the Level 1 problem even under simultaneous conditions were replaced, as were seven *S*s who failed to follow instructions, such as those requiring *S*s to give a guess after each instance.

After each *S* was given a practice and experimental booklet *E* read aloud the three pages of orientation material while *S*s read the same material silently. In this material *S* was exposed to the stimulus universe and was shown a number of instances of concepts involving a single value each (practice-level problems) and was introduced to the role of negative instances as they are used in presenting information pertinent to simple concept problems. The *S*s were then asked to solve three practice-level problems. They typically accomplished this in 2 min. or less, thus indicating their assimilation of the explanatory materials given them to that point. Subsequently *S*s were practiced on problems of Levels 1 and 2. Since all of the relevant experimental materials were contained in booklets it was possible to have each *S* work on a different problem during the same experimental session.

Following the practice session *E* emphasized the importance of the standardized test conditions which required *S*s to turn the pages in their test booklets exactly on cue at the end of each 30-sec. period. Specifically 20 sec. were allowed for perusal of the instances and up to 10 sec. for recording of tentative concept hypotheses. No questions were permitted during the learning-testing phase and *S*s were given their four test problems consecutively with a rest period of about 2 min. intervening between problems.

RESULTS

The extent to which memory plays an important role in concept learning is shown by the comparison of the number of correct statements of the concepts after the presentation of the minimum number of instances (after Trial 10 for Level 1 problems and after Trial 4 for Level 2 problems) under the simultaneous and successive conditions. Eleven test trials were involved in the former series (after Instances 10 through 20) and five in the latter (after Instances 4 through 8). Success scores determined on this basis were not normally distributed and tests of significance of success as a function of type of stimulus presentation and problem complexity level were made by the Wilcoxon (1945) nonparametric test for paired-replicates. The results of these tests, as well as the median number of correct responses, for the different stimulus presentation conditions are summarized in Table 1. Here it may be seen that concept attainment under simultaneous presentation is markedly superior at both levels of problem difficulty ($P < .001$ for Level 1 and $P = .02$ for Level 2). The findings replicate those of Hovland and Weiss (1953) in showing poorer performance when previous instances must be retained in memory rather than being directly available for inspection.

TABLE 1

NUMBER OF CORRECT STATEMENTS OF CONCEPT AS A FUNCTION OF TYPE OF STIMULUS PRESENTATION AND PROBLEM

Condition	No. Nonzero Differences	Median	Diff.	One-Tail P
Level 1 problem				
Simultaneous	34	10.8	10.4	.001
Successive		0.4		
Level 2 problem				
Simultaneous	28	4.6	4.2	.02
Successive		0.4		

Note.—Level 1 problem involved instances 10 through 20, with maximum score of 11; Level 2 problem involved instances 4 through 8, with maximum score of 5.

Errors were analyzed to determine how many were responses incompatible with the immediately presented instance and how many were incompatible with instances presented somewhere earlier in the series. The former may be described as "perceptual-inference" errors since they are responses to stimuli physically available and constitute failure either to perceive correctly the stimulus presented or to make the appropriate inference as to possible concepts from it.

Perceptual-inference errors were reltively infrequent. With the Level 1 problem 5 *S*s made a total of 6 errors under simultaneous conditions and 7 *S*s made a total of 8 errors under successive conditions. With the Level 2 problem 12 *S*s made a total of 24 errors under simultaneous conditions and 15 *S*s made a total of 21 under successive. As would be expected, no reliable difference between simultaneous and successive conditions was obtained for the just-presented instance. The difference in number of errors at the two levels of difficulty is, however, significant ($P < .01$).

The *S*s who made these perceptual-inference errors not unexpectedly had greater difficulty in learning the concepts. These results are reported in Table 2. The *S*s were divided into "high" and "low" groups in terms of the number of "perceptual-inference" errors made. The significance test is based on the Mann-Whitney *U* technique (1947). The number of *S*s in the high and low groups differ among the various conditions since *S*s could not sensibly be divided at the median number of errors. It can be seen that the differences were greater at the greater complexity level and in the simultaneous as opposed to the successive presentation conditions. Only in the sucessive conditions at Level 1 did the difference fail to approach at least marginal significance. Particularly interesting also is the fact that a difference approaching significance was obtained in the simultaneous condition at Level 1 where only five *S*s committed perceptual-inference errors. Thus it appears that at even the simplest level the task of formulating a concept compatible with the immediately presented instance exhibits the operation of wide individual differences which significantly affect over-all performance.

The other type of error mentioned above involves the stating of a concept hypothesis which is compatible with the just-presented instance but fails to take into account information received on an earlier trial. These "memory" errors are quite numerous and were analyzed in detail. They likewise are highly correlated with success in attaining the correct concept. The *S*s are divided into "high" and "low" groups in terms of the number of memory errors made. Significance tests are again based on the Mann-Whitney *U* technique.

An analysis was made of the number of concept hypotheses which were incompatible with information on instances varying numbers of trials earlier. For example, on Trial 7 an *S* gives an hypothesis as to the correct concept. This may or may not be consonant with the instance just presented. One can also check to see whether it is or is not consonant

TABLE 2

SUCCESS IN CONCEPT ATTAINMENT AS RELATED TO NUMBER OF PERCEPTUAL-INFERENCE ERRORS

Condition	Level	Perceptual-Inference Performance						P*
		High			Low			
		N	Median Perceptual-Inference Errors	Median Correct Answers	N	Median Perceptual-Inference Errors	Median Correct Answers	
Simultaneous	1	31	0	10.8	5	1.1	8.0	.06
Successive	1	29	0	3.8	7	1.1	0.4	.12
Simultaneous	2	24	0	10.9	12	2.0	0.4	<.001
Successive	2	21	0	7.0	15	1.1	0.1	.02

* One-tail difference in number of correct answers.

TABLE 3

SUCCESS IN CONCEPT ATTAINMENT AS RELATED TO MEMORY CAPACITY

Condition	Level	Memory Capacity						p*
		High			Low			
		N	Median Memory Errors	Median Correct Answers	N	Median Memory Errors	Median Correct Answers	
Simultaneous	1	24	0.0	11.0	12	2.2	8.0	<.001
Successive	1	17	5.7	7.0	19	11.5	0.0	<.001
Simultaneous	2	19	0.0	11.0	17	2.4	0.4	<.001
Successive	2	17	0.3	10.8	19	5.4	0.0	<.001

* One-tail difference in number of correct answers.

with the instance presented one trial earlier, two trials earlier, etc. In the case of the simultaneous series one has physically available the earlier instances and differential failure to utilize earlier instances is primarily attributable to failure to consider the instances presented earlier. But in the case of the successive presentation incompatibility is also attributable to failure to remember the earlier instances and/or to remember their implication. It is thus a sensitive measure of memory. The conditions for the two series are otherwise identical so that the simultaneous series constitutes a "control" condition for the successive. The Level 1 problem provides the most extensive data since a longer series of instances is involved than is the case with the Level 2 problem. The number of errors was calculated by determining the number of occasions in which a response was given which was incompatible with the implications of instances varying numbers of trials earlier. Since the distribution of *S*s' errors was J-shaped, medians are employed. The number of errors for each degree of remoteness from the immediately presented instance are given in Table 4. Also presented are the number of opportunities for making an error. The variation in number of opportunities is attributable to the fact that there is some redundancy inevitable in the series so that sometimes a concept hypothesis is excluded more than once. Such cases are excluded from the analysis with the result that there are fewer opportunities to make errors as the degree of remoteness increases. While it would have been possible to compute errors up to 19 degrees of remoteness, the analysis was restricted to the range in which at least 50% of the initial number of opportunities was permitted. This gave a range from

TABLE 4

MEDIAN NUMBER OF ERRORS: RESPONSES INCOMPATIBLE WITH INSTANCES PRESENTED VARYING NUMBERS OF TRIALS EARLIER

Condition	Degree of Remoteness							
	0	1	2	3	4	5	6	7
Successive	.03	.14	.25	.40	.36	.66	.67	.74
Simultaneous	.06	.12	.03	.05	.03	.00	.03	.02
Concepts excluded[a]	78	78	68	65	57	55	50	45

[a] Number of concepts excluded for the first time by the instance being considered.

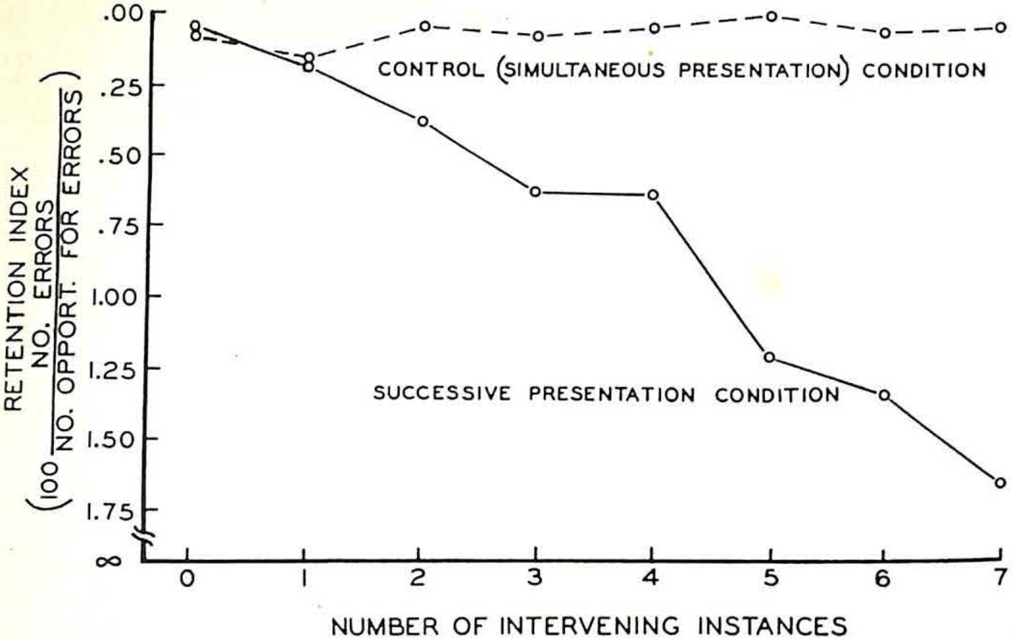


FIG. 1. Retention index for information after varying numbers of intervening instances.

zero to seven degrees of remoteness. For presentation of the decrease in utilization of the information contained in the instances more remote in the series it seemed appropriate to use a ratio of errors to number of opportunities. Accordingly, in the graph presented in Fig. 1 the median number of errors is divided by the number of opportunities for making such an error. This fraction is then multiplied by 100. The higher the value of this index, the greater the extent of "forgetting." The solid line indicates the changes in ratio with degree of remoteness between the concept given and the instance relating to that hypothesis for the successive conditions, in which memory for past instances was required. Results for the simultaneous presentation condition provide a good control for any possible contamination or selection; they are given by the broken line. Clear evidence of increased "forgetting" with increased remoteness is obtained. The divergence in slope between the two conditions is highly significant ($P < .001$). Comparable results are found for the Level 2 problem, but only three points can be utilized.

Analysis was also made of the extent to which information presented early in the series is correctly utilized as compared with that presented late, for instances of the same degree of remoteness. Figure 2, in which the abscissa and ordinate are in the same units used in Fig. 1, shows the memory effects for instances presented early and late in the series. The broken line represents the error ratio for concepts excluded early (Trials 1–3) and the solid line portrays the ratio for concepts excluded late (Trials 7–10). (To make the number of opportunities for error similar, four instances were used for the later series and three for the earlier series.) The results indicate the operation of a moderately strong "primacy effect," i.e., the information presented early in the series is utilized more effectively than the information presented later on. The difference in slope is significant ($P < .01$). Both primacy and recency effects can be said to be operative—primacy as shown by the higher level of retention of the early series of instances, and recency as shown by the fact that the curves decline with increased degree of remoteness for both the early and late series.

DISCUSSION

The present study indicates that a very low proportion of errors in concept formation of the type studied here is attributable to failure to assimilate the information perceptually available and to draw the appropriate inferences concerning the concept being sought. More of the errors are attributable to failure to remember earlier instances in such a way as to see their implications. These errors increase progressively with increasing numbers of intervening instances. The decline in retention, the "forgetting curve," is approximately linear over the range examined.

These data indicate a close tie between phenomena of concept acquisition and those of memory (cf. also Underwood, 1952). Further studies may profitably be devoted to exploring the overlap between these areas, by varying the similarity of materials and extending the range of time intervals.

For machine simulation of concept learning some explicit treatment of memory effects is necessary. Up to the present it has been simpler to assume unlimited memory, as in the model developed by Hovland (1952) where the implications of different types of instances is explored when *S* knows the nature of the concept task and is able to retain all of the information presented. For realistic simulation of more typical concept learning, however, we must cover also situations where the knowledge of the type of concept is quite incomplete ("open") and also where there is typical human fallibility of memory.

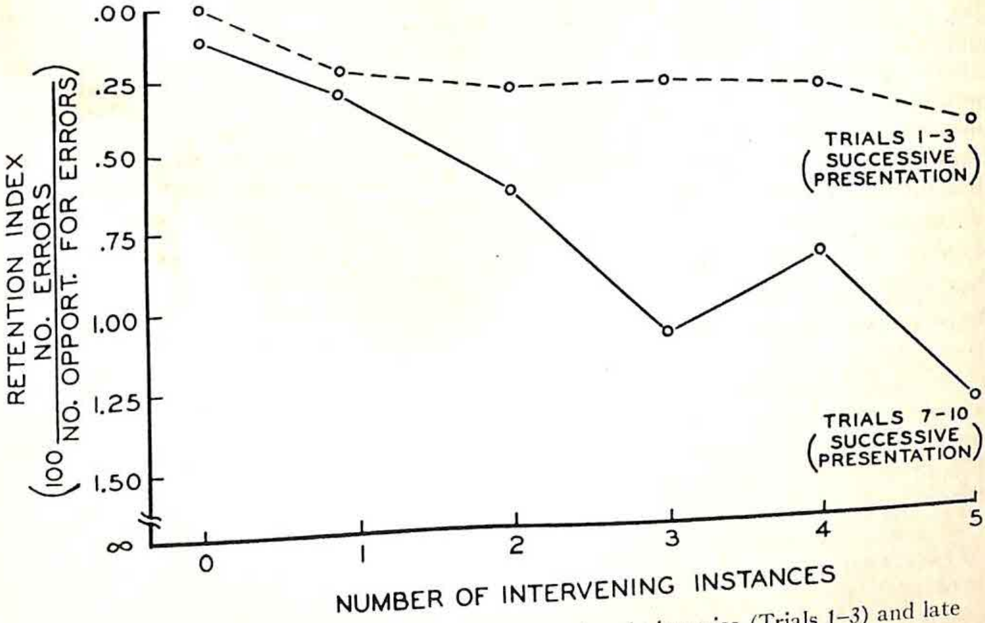


FIG. 2. Retention indices for information presented early in series (Trials 1–3) and late in series (Trials 7–10) after varying numbers of intervening instances.

The intriguing task is to build this forgetting into the machine in the most effective manner. It would be relatively simple to do this on an empirical basis, following known results for the function. What is more challenging is to build the memory function into the computer in ways which best meet theoretical requirements concerning interference effects as a basis for the loss. Work on this formulation for simulation is currently underway in our laboratory.

The type of "primacy effect" obtained in this experiment deserves further investigation. Two types of explanations suggest themselves for further exploration. It may be that material presented early is attended to more closely and is in some way more prominent and impressive. Or it may be that at the outset there are fewer instances which must be considered in drawing inferences as to the appropriate concept while later more instances have to be considered and remembered, and this leads to greater difficulty.

The generality of the present findings on memory is not known. Replication is needed with different types of concept learning material, both verbal and nonverbal. In addition, studies are needed which use positive instances, and combinations of positive and negative instances, to supplement the results in the present experiment where only negative instances were employed. Memory effects should show up most clearly in the difficult all-negative series.

SUMMARY

Series of negative instances were used to define various concepts. After each instance *S* was instructed to guess the characteristics of the concept. Under experimental conditions each instance was removed from view when the next instance was presented. Under control conditions all prior instances remained in view.

Memory effects were studied by comparing the number of cases where the guess was incompatible with information presented in prior instances under these two conditions. Guesses were seldom incompatible with the instance just presented but under the experimental conditions were increasingly discrepant from instances further removed, giving rise to a characteristic "forgetting curve." The rate of forgetting was less rapid for the instances presented early in the total series than for those presented late. Significant individual differences were obtained, indicating that *S*s who made more "perceptual-inference" and "memory" errors had greater difficulty in acquiring the concepts.

REFERENCES

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **53**, 461–472.

HOVLAND, C. I. Computer simulation in the behavioral sciences. Talk given on *Voice of America*, 1959. (Mimeo.)

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

UNDERWOOD, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59**, 209–220.

WILCOXON, F. Individual comparisons by ranking methods. *Biomet. Bull.*, 1945, **1**, 80–82.

(Early publication received
October 21, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

THE RELATIONSHIP OF INDUCED MUSCULAR TENSION, TENSION LEVEL, AND MANIFEST ANXIETY IN LEARNING[1]

O. IVAR LOVAAS

University of Washington

This study investigated the relationship between manifest anxiety (MAS), individual differences in muscular tension (MT), and induced muscular tension (IMT) in a learning task. The *S*s were split into a high- and low-scoring group on MAS, and a high- and low-scoring group in terms of spontaneous eyeblink rate (MT). Meyer (1953) has argued for the efficacy of the eyeblink rate as a measure of muscular tension. The performance of these groups could be observed under various IMT conditions. Paired-associate learning was employed. Adjectives were paired in such a manner as to produce pairs of high and low degrees of association and competing pairs, the latter being pairs where two or more response words were associated with the stimulus word.

The rationale for this study was based on previous empirical investigations which have indicated a close relationship between muscular tension and anxiety. Malmo (1957) reported that people adjudged anxious by clinicians obtained high scores on indices of MT. Lovaas (1960) obtained data showing that IMT and MAS have similar effects upon performance in a learning task. Meyer (1953) conceptualized the energizing properties of drive to be muscular tension.

Spence (1956) has presented a theoretical framework and empirical observations which imply that *S*s who are differentiated on MAS also are differentiated on *D*; he also related *D* to performance in paired-associate learning. In view of the close relationship of IMT and MT to MAS, this study sought to investigate the possibility whether all three variables could be related to each other by Hull's generalized drive and whether they would be supported as *D* variables in paired-associate learning. If MAS, MT, and IMT were to be supported as *D* variables, then (*a*) the joint occurrence of MAS and IMT and of MT and IMT would summate in their effect upon performance; (*b*) performance should change in the same manner with increasing MT as with increasing IMT; and (*c*) increasing magnitude of IMT, MAS, and MT, singly or in combination, should be more beneficial to the highly associated pairs than to the competing pairs, with the effect upon the lowly associated pairs falling between these extremes.

METHOD

Subjects and apparatus. The short 20-item Bendig (1956) revision of the Taylor Manifest Anxiety Scale was administered to 110 male, undergraduate students enrolled in introductory psychology courses. The Taylor items were embedded in items from the *K* scale of the MMPI. All students with a raw *K* score of 24 or more were eliminated from further consideration. This procedure eliminated 12 students from the extreme low end of the distribution. Of the remaining

---

[1] This article is taken from a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Washington, 1958. The author is indebted to Moncrieff H. Smith, Jr., under whose direction the thesis was prepared.

TABLE 1

SETS OF PAIRED ASSOCIATES

Set 1:		Set 2:	
Stimulus	Response	Stimulus	Response
Adept	Gallant	Adept	Gallant
Complete	Thorough	Empty	Vacant
Distant	Bashful	Mammoth	Wanting
Empty	Vacant	Wicked	Evil
Frigid	Balmy	Complete	Gentle
Insane	Crazy	Frigid	Heavy
Mammoth	Wanting	Pious	Tired
Pious	Devout	Urgent	Wholesome
Urgent	Extinct	Barren	Grouchy
Wicked	Evil	Little	Headstrong
		Roving	Double
		Tranquil	Leading

98 students the 32 lowest scoring students were selected for the low-MAS *S*s ($M = 1.9$, range = 0–4), the 32 students who scored highest on the scale were selected for the high-MAS *S*s ($M = 8.2$, range = 5–17). None of the *S*s had experience with the kind of learning task presented. They were unaware of the basis for selection and of the purpose of the experiment.

Each pair of words was typed on a card and the cards were mounted on two screws. As the screws rotated they brought the cards forward to an exposure window and dropped one card every 4 sec. A shutter mechanism allowed the stimulus word to be shown for 1.7 sec. of the 4-sec. interval, then exposed the response word for 2.3 sec. Inadequacies in the apparatus allowed an *SD* of about .1 sec. in the various exposure times.

A Smedley dynamometer was located on each side of the exposure window, about 25 in. apart and about 20 in. from *S*. The handles were padded. The *S*s alternated hands from one trial to another. The pointer on the dynamometer closed a circuit which lighted a bulb, located 2 in. above one corner of the exposure window, whenever the pressure exerted by *S* was ±3 lb. of the pressure required of him.

Material and procedure. On arriving for the experiment each *S* squeezed the dynamometers, one at a time, as hard as he could for 30 sec. each. Except in the zero IMT condition, *S* would squeeze at a designated fraction ($\frac{1}{8}$, $\frac{1}{4}$ or $\frac{1}{2}$) of the maximal pressure exerted during the 30th sec. on the respective handle. The tension signal light was kept on by *S* during each trial. The *S*s in the zero IMT condition merely gripped the handles. The contact on the dynamometer allowed a pressure of only 3 lb. to turn the light off. This condition was introduced to equalize distracting effects from the light on the groups.

The *S* was told that he would work for a predetermined number of trials, that if he got all correct at one time he should attempt to get all correct again, and that he should do his best on each trial. The instructions were designed to avoid the incentive motivation effects of the tension conditions which might have operated if *S*s knew that faster work would accomplish earlier release.

The *S* was given a practice session of six trials on 10 paired nouns, with the use of the dynamometers begun on Practice Trial 3. A 1-min. rest period followed the practice trials after which two sets of paired adjectives were learned (see Table 1). All *S*s learned Set 1 first, then Set 2.

Using Haagen's (1949) word list, five "H pairs" were selected for Set 1 in such a manner as to maximize strength of association between the paired stimulus and response words. The remaining five pairs "L pairs" were selected in such a manner that the stimulus word was paired with a response word with which it had little associative connection. In addition, the pairs were selected to minimize association value between one pair and another. No beginning letter or suffix was repeated either in the stimulus words or in the response words. Set 2 was composed in the following manner. For the "H pairs" four pairs were the same as those used in Set 1. These had high $_{S}H_{R}$ since *S*s had had a number of trials on these pairs immediately prior to being presented with Set 2. For the four "C pairs," the stimulus words were the same as in Set 1 but the response words were changed, thus creating many competing responses. The "L pairs" were four altogether new pairs selected to have low associative value between the stimulus and response words, in relationship to the "H pairs." As with Set 1, the pairs on Set 2 were selected to minimize association between one pair and another.

The two MAS groups were split into four subgroups having different levels of induced muscular tension: control group, which merely gripped the handles; and groups with $\frac{1}{8}$, $\frac{1}{4}$, and $\frac{1}{2}$ pressure on the hand dynamometer. For each condition on Set 2 there were two *S*s from each of the conditions on Set 1, a total of eight *S*s for each of the conditions. This assignment of *S*s provided for the testing of possible interaction of pressure on Set 1 with performance on Set 2.

Set 1 was presented for a minimum of two consecutive errorless trials or for a maximum of 21 trials. There was a 4-min. interval between presentation of Sets 1 and 2 during which *S* rested and *E* changed the contacts on the dynamometer if *S* had been assigned to a different pressure condition on Set 2. Instructions for Set 2 included a statement to the effect that some of the pairs were the same as in Set 1 and some were different. Set 2 was presented for a maximum number of 14 trials. The pairs were shuffled after each presentation to avoid serial effects. Intervals between the trials were held constant at 10 sec. The *Ss'* performance on the learning task was recorded in terms of number of correct anticipations for each pair. All correct anticipations were pronounced during the 1.7-sec. anticipation interval.

Number of eyeblinks was counted on sight by *E* on Trial 2 of the practice set, and on Trials 1, 11, and 21 (or on the trial following two consecutive errorless trials) on Set 1. Only complete closures of the eyelid were counted. The *E* was aware of *S*'s classification with respect to both MAS and IMT at the time the eyeblinks were counted.

RESULTS

The relationship of MAS and IMT. Means and *SD*s for correct anticipations on pairs in Sets 1 and 2 for the high- and low-MAS groups under the IMT conditions are presented in Table 2. The analysis of variances on these data shows that only the effects of variation in kinds of pairs were significant ($F = 247.1$ on Set 1; $F = 21.1$ on Set 2; both have $P < .01$). There was no support for a summative effect of MAS and IMT upon performance, or for a differential effect of the magnitude of these variables, singly or in combination, upon kind of pair. There was no effect of pressure on Set 1 on the performance on Set 2.

TABLE 2

MEANS AND *SD*s OF CORRECT ANTICIPATIONS FOR HIGH- AND LOW-MAS *Ss* UNDER IMT ON DIFFERENT PAIRS IN SET 1 AND 2

Set 1

MAS	High Association				Low Association			
	IMT				IMT			
	0	⅛	¼	½	0	⅛	⅛	½
High	*M* 75.4 *SD* 15.3	77.5 15.2	79.6 10.6	80.5 7.5	45.8 13.6	46.8 24.1	54.0 23.5	43.5 19.2
Low	*M* 83.0 *SD* 7.5	82.9 6.7	80.6 10.4	72.8 13.0	57.5 24.4	58.8 18.3	44.0 23.3	40.9 22.2

Set 2

MAS	High Association				Low Association				Competing Association			
	IMT				IMT				IMT			
	0	⅛	¼	½	0	⅛	¼	½	0	⅛	¼	½
High	*M* 44.2 *SD* 5.6	42.0 8.8	41.9 8.8	38.0 4.8	36.4 9.6	30.9 9.7	28.2 14.0	28.9 13.4	26.1 11.5	17.9 14.1	18.0 9.6	12.1 9.5
Low	*M* 41.9 *SD* 11.0	45.9 8.1	45.1 4.8	36.1 12.8	32.2 9.3	32.4 12.8	34.8 4.9	23.2 11.7	22.8 18.2	19.4 16.7	23.8 9.7	14.9 11.9

TABLE 3

ANALYSIS OF VARIANCE FOR GROUPS HIGH AND LOW ON EYEBLINK RATE AND INDUCED MUSCULAR TENSION ON SET 2

Source of Variation	*df*	*MS*	*F*
High Associated Pairs			
Eyeblink × Pressure	3	162.53	2.528*
Within	36	64.29	
Total	43		
Low Associated Pairs			
Eyeblink × Pressure	3	280.73	2.766*
Within	36	101.50	
Total	43		
Competing Pairs			
Eyeblink × Pressure	3	489.28	3.515**
Within	36	139.19	
Total	43		

Note.—This analysis follows the outline that Snedecor (1946) has given for analysis of variance in the case of disproportionate subclass numbers.
* $P < .06$.
** $P < .05$.

The relationship of MT and IMT. To test the hypothesis of a summative effect of MT and IMT upon performance, *S*s from the high- and low-MAS groups were matched on basis of their eyeblink rate during Trial 2 of practice. Twenty-two pairs were matched, seven of which were one blink apart, two were two blinks apart, one was five blinks apart, and the remaining pairs had equal blink rates. These 44 *S*s were split into a low-MT group, those having 12 or fewer blinks; and a high-MT group, those having 13 or more blinks. The mean blink rate of the low-MT group was 7.3 (7.5 for low-MAS *S*s and 7.2 for high-MAS *S*s) with an *N* of 24. The mean blink rate of the high-MT group was 21.1 (21.6 for the low-MAS *S*s and 20.8 for the high-MAS *S*s) with an *N* of 20. The *S*s were selected to assure no fewer than four *S*s for each IMT condition on Set 2.

Table 3 gives the results of the analysis of variance for performance on Set 2 of the high- and low-MT groups under the IMT conditions. The interaction between low- and high-MT groups under various degrees of IMT was significant on the competing pairs ($P < .05$). The *F* associated with the same interactions on the two other kinds of pairs had approximately a $P < .06$. Although this did not meet the statistical criteria for significance, these results might be considered to have reflected a reliable difference because the small number of *S*s would leave a relatively large error variance and because the means were arranged in the predicted direction. The results of this analysis are presented graphically in Fig. 1.

As can be observed the results were consistent with the summative effect of MT and IMT upon performance. High-MT *S*s tended to perform better than low-MT *S*s under 0 and $\frac{1}{8}$ IMT conditions, with reversal of this differ-

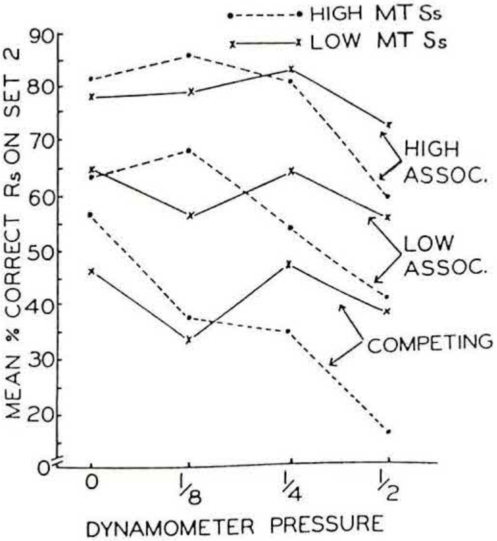


FIG. 1. Paired-associates learning of high- and low-MT (eyeblink rate) *S*s, matched on MAS (Manifest Anxiety), under various conditions of induced muscular pressure on Set 2. Word pairs of high and low association value and competing pairs were interspersed within the same set, but were analyzed separately.

ence under $\frac{1}{4}$ and $\frac{1}{2}$ IMT conditions. Under maximal IMT the performance of high-MT *S*s was consistently inferior to the performance of low-MT *S*s. The statistical design (Table 3), provided no test for the differential effect of increase in MT or IMT upon kind of pair. However, inspection of Fig. 1 shows the data do not support this hypothesis. Although the curve forms for the high-MT *S*s were consistent with this hypothesis (the level of IMT associated with optimal performance was higher for the "L and H pairs" than for the "C pairs"), the curve forms for low-MT *S*s did not reflect this interaction of kinds of pairs and IMT in that the optimal level of IMT was independent of kinds of pairs.

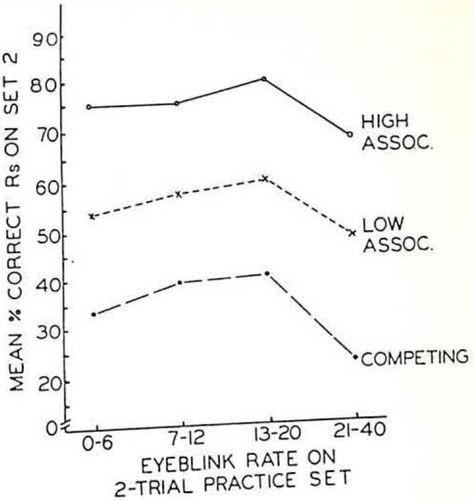


FIG. 2. Paired-associate learning of various MT (eyeblink rate) groups. Word pairs of high and low association value and competing pairs were interspersed within the same set, but were analyzed separately.

The relationship of MT to performance. It was hypothesized that if MT and IMT were separate *D* variables performance should change in the same manner with increasing MT as with increasing IMT. To test this hypothesis *S*s were split into four groups on the basis of their eyeblink rates on Trial 2 of practice. Each group had no fewer than 14 and no more than 18 *S*s. This grouping tended to maximize the *N* of each group and give the same number of points (four) on which to observe the curve relating MT to performance as was the case with the curve relating IMT to performance. Figure 2 presents a graphic description of the relationship of eyeblink rate (MT) to performance on the various pairs of Set 2. This relationship is similar to that based on the performance of *S*s working under increasing IMT on lists of nonsense syllables (Courts, 1942). Discrepancies between the curves in Fig. 1 and 2 might be due in part to error variation based on small number of *S*s ($N = 8$) for each group in Fig. 1. This opinion is based on the fact that in most research where statistically significant effects have been obtained on the relationship of IMT to performance, the curve forms are either continually increasing or decreasing in relationship to some level of IMT associated with optimal performance.

Additional findings on the eyeblink rate. The reliability of the eyeblink rate could be approximated only because it was not observed twice under strictly comparable conditions. By correlating the counts on Trial 2 of practice, where there was no IMT for any *S*, with the counts on Trial 1 on Set 1, which included all IMT conditions, a product-moment coefficient of .76 was obtained ($df = 62$; $P < .01$).

The relationship of eyeblink rate (from Trial 2 on practice) to MAS scores was curvilinear, with extremely low and high scores on the MAS related to high blink rate. MAS scores around the mean were associated with the lowest mean blink rate

TABLE 4

Eyeblink Rates Associated with Variations in Manifest Anxiety Scale Scores Data from Trial 2, Practice Set

Eyeblink Rate	Manifest Anxiety Scale Intervals					
	0–1	2	3–4	5–6	7–10	11–17
N	14	14	13	16	12	7
Mean	14.5	11.1	9.9	14.5	14.5	19.7
SD	10.4	8.7	7.9	6.3	9.6	9.5

and also with the lowest variability in blink rate (Table 4). Eta between MAS scores and blink rate yielded a coefficient of .37, this relationship being based upon an *N* of 76. There was a significant difference in blink rate between the high-MAS and low-MAS groups ($F = 7.06$; $P < .01$).

There was a tendency for the eyeblink rate to increase after Trial 1 (Set 1), then decrease somewhat after mastery was reached.

In the discussion so far, eyeblink rate has been considered as a response which reflected underlying generalized muscular tension. Considering the eyeblink as a response per se, response related to variation in IMT could be observed. Fig. 3 graphically pictures the eyeblink rates from Trial 1 on Set 1 and shows the effect of various degrees of IMT on this response rate.

The MAS × IMT interaction on these data does not quite reach significance ($F = 2.44$; $df = 3$ and 56; $P <$ approximately .06) but the interaction is in the predicted direction. These data are consistent with the summative effect of MAS and IMT in that the level of IMT associated with the highest response rate is lower for high-MAS than for low-MAS *S*s. It is to be noted that this was the only instance in this study where high-MAS *S*s had a higher response rate than low-MAS *S*s under the zero IMT condition, thereby reflecting their hypothesized higher level of *D*.

Discussion

Relationship of MAS, MT, and IMT to each other. It was hypothesized that the joint occurrence of IMT and MAS would summate in their effect upon performance. Data from this study did not confirm this hypothesis. After the data had been collected and analyzed, a study by Meyer and Noble (1958) was published. Their hypothesis concerning the relationship of IMT and MAS, similar to the one for this study, was supported by significant interaction between MAS and IMT. Low-MAS *S*s tended to improve while high-MAS *S*s tended to deteriorate with IMT. Their procedure was slightly different from the one in this study in that they used only one IMT condition, employed 20 *S*s in each of their groups, and had a verbal maze for their learning task. On the basis of the larger *N* in their study it seems preferable to accept their results rather than those from this study.

The hypothesis was confirmed that IMT and MT would summate in their effect upon performance. Thus the level of IMT associated with optimal performance was smaller for high-MT

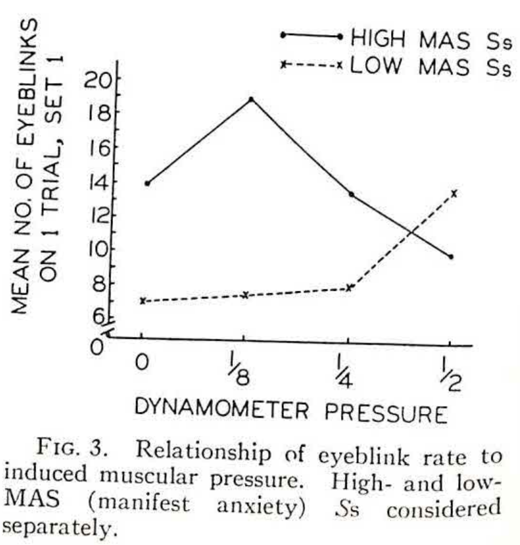


Fig. 3. Relationship of eyeblink rate to induced muscular pressure. High- and low-MAS (manifest anxiety) *S*s considered separately.

*S*s than for low-MT *S*s. This formulation is similar to Stauffacher's (1937). In brief, his hypothesis stated that poor performers function on a lower level of "normal" tension than good performers and they therefore benefit more than the good ones from IMT. However, Courts (1939) did not find that the optimal dynamometer tension occurred earlier for the good performers than for the poor ones and concluded that the differential effect is slight, if it exists. In a second part of the study, Courts recorded the amplitude of the knee-jerk during the memorization trials for his *S*s, employing this measure in the same way eyeblink rate was used in this study. Subsequently he separated out the *S*s who showed the greatest versus the lowest gain from IMT and compared the amplitude of their knee-jerks. He found that *S*s who failed to benefit from IMT tended to show the greater increase in knee-jerk amplitude during memorization. The latter findings are consistent with the data of the present study.

The data on the relationship of MT and IMT also show that MT and IMT are "interchangeable" variables with respect to the performance observed in this study. Previous research (Lovaas, 1960) has indicated that IMT and MAS may also be "interchangeable." Performance in a learning situation of high-MAS *S*s was closely approximated by *S*s assumed to score on a lower level of MAS working under IMT.

In view of the fact that the high-MAS *S*s emitted a higher eyeblink rate than low-MAS *S*s, it is suggested caution be used in accepting conclusions based on eyelid conditioning about the more rapid conditioning of high-MAS versus low-MAS *S*s (Spence, 1956). It is questionable whether this problem could be circumvented by employing high- and low-MAS *S*s matched on initial eyeblink rate since differentiations on *D* may carry differentiations on eyeblink rate with them.

Relationship of MAS, MT, and IMT to performance. The fact that the differentially associated pairs employed in this study did not discriminate between different magnitudes of MAS, MT, and IMT is inconsistent with Spence's (1956) data on MAS and Lovaas' (1960) data on IMT. The learning task employed perhaps was not sufficiently sensitive. The *E* observed a tendency in some *S*s to "withhold" the response word for the "H pairs" in Set 2 until they knew which pairs had been changed and which remained unchanged. Any differential effect of tension on the various pairs might then be considered to be more significant than the effect allowed by the statistical design.

Previous research has been concerned with possible changes in the nature of the learning task as learning progresses and the concomitant change in amount of IMT associated with optimal performance. It has been found that the level of IMT associated with optimal performance decreases as learning progresses (Courts, 1942). The smaller magnitude of IMT as learning progresses might reflect merely the increasing magnitude of MT and not bear any relationship to change in the nature of the learning task. This seems possible in view of the findings from this study on the summative effect of MT and IMT and in view of the fact that MT increases after the beginning of acquisition. Data in this study showed that the eyeblink rate increased after the first trial, a finding similar to that of Bartoshuk's (1955) on the EMG. Changes in tension may occur as a function of requirements put upon *S* by both task and *E*. For example, Stennet (1957) found that motivational instructions increased the muscle tension of his *S*s. If IMT does not facilitate or even impairs performance on "difficult" or "complex" tasks, it might well be that these tasks induce muscle tension in *S*. Adding hand dynamometer tension would bring the level of tension beyond the level associated with optimal performance. In investigations concerned with the effect of IMT on performance, caution should be exercised not to contaminate the effect of this variable with uncontrolled variation in MT.

SUMMARY

Subjects scoring high and low on the Taylor Manifest Anxiety Scale (MAS) worked under various degrees of induced muscular tension (IMT) in paired-associate learning. The *Ss* also were split into a high and low group on the basis of their eyeblink rates, assumed to be an index of individual differences in muscular tension (MT). The learning task, paired adjectives, was constructed with the intent of producing highly associated pairs, lowly associated pairs, and pairs with many competing responses.

It was hypothesized that MAS and IMT, as well as MT and IMT, were related to each other by Hull's generalized drive. Thus MAS and IMT, and MT and IMT, should summate and be interchangeable in their effect on performance. The three variables were related, as Spence has related *D*, to performance in a paired-associate learning task. Hence, increasing magnitude of IMT, MAS, and MT, singly or in combination, should be more beneficial to the highly associated pairs than to the competing pairs, with the effect upon the lowly associated pairs falling between these extremes. While there is support for a summative and interchangeable relationship between MT and IMT, there is no support for a summative relationship between MAS and IMT, or for a differential effect of the magnitude of any of these variables upon the various pairs.

Other findings included a curvilinear relationship between eyeblink rate (MT) and MAS, but high scores on MAS were associated with the highest blink rate.

REFERENCES

BARTOSHUK, A. K. Electromyographic gradients in goal-directed activity. *Canad. J. Psychol.*, 1955, **9**, 21–28.

BENDIG, A. W. The development of a short form of the manifest anxiety scale. *J. consult. Psychol.*, 1956, **20**, 384–385.

COURTS, F. A. Relations between experimentally induced muscular tension and memorization. *J. exp. Psychol.*, 1939, **25**, 235–256.

COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347–367.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

LOVAAS, O. I. Supplementary report: The relationship of induced muscular tension to manifest anxiety in learning. *J. exp. Psychol.*, 1960, **59**, 205–206.

MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–287.

MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.

MEYER, D. R., & NOBLE, M. E. Summation of manifest anxiety and muscular tension. *J. exp. Psychol.*, 1958, **55**, 599–602.

SNEDECOR, G. W. *Statistical methods.* Ames: Iowa State Coll. Press, 1946.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

STAUFFACHER, J. C. The effect of induced muscular tension upon various phases of the learning process. *J. exp. Psychol.*, 1937, **21**, 26–46.

STENNET, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–62.

(Received January 31, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

STUDIES IN INCIDENTAL LEARNING: VIII. THE EFFECTS OF CONTEXTUAL DETERMINATION[1]

LEO POSTMAN AND PAULINE AUSTIN ADAMS

University of California, Berkeley

Earlier studies in this series have supported the conclusion that intent to learn significantly influences *S*s' behavior during exposure to the learning materials. What features of the materials *S*s discriminate and how they respond to individual items and sequences of items depends on whether or not they anticipate a test of retention. Specifically, incidental *S*s are more selective than intentional *S*s, i.e., they make fewer differential responses to the learning items. Such responses, we assume, are essential for the differentiation of the materials and also serve to mediate the reproduction of the items at the time of the retention test. There appear to be two major ways in which instructions to learn influence *S*s' differential responses to the learning materials. First, incidental learners respond selectively to individual items which are relatively familiar or meaningful. Intent to learn serves to extend the range of items which elicit differential responses. Thus, the higher the association value of the items the smaller is the difference between the amounts recalled by intentional and incidental learners (Postman, Adams, & Phillips, 1955). Second, intentional learners are more likely than incidental learners to respond to features of the materials which facilitate linkages among the members of a series. Intentional learners surpass incidental learners in the recall for serial order even when the instructions make no reference to the order of the items (Postman, Adams, & Bohm, 1956).

These findings lead to certain expectations concerning the effects on intentional and incidental learning of varying degrees of contextual determination of the materials. Degree of contextual determination refers to the extent to which the choice of each successive item in a series depends on the items preceding it (Miller & Selfridge, 1950). Two effects of verbal context on retention may be distinguished. First, the presence of sequential dependencies favors intraserial associations, and hence the orderly reproduction of the members of the series. Since instructions to learn increase sensitivity to serial order, sequential dependencies among the items in a list should be more beneficial to serial learning under intentional than under incidental conditions. Second, within a given series, the degree of dependent probability between successive items will usually not be uniform throughout but will vary around some mean value. The more strongly an individual item in such a series is determined by the verbal context the more likely it is to be recalled when that context is reinstated at the time of the retention test. The results of the experiments on association value suggest that the effects of verbal context on the differential recall of individual items should be at least as great under incidental as under intentional conditions. Thus, when the test of retention calls for the reproduction of a series of items,

---

[1] This research was facilitated by a grant from the Behavioral Sciences Division of the Ford Foundation.

reductions in the uncertainty of the serial order should produce greater increases in recall for intentional than for incidental learners. Selective recall of specific items strongly determined by the verbal context will, however, occur under both conditions.

The present study reports two experiments in which the effects of verbal context on intentional and incidental learning were evaluated. Two methods of varying degree of contextual determination were used. In Exp. I intentional and incidental *S*s were exposed to lists representing different degrees of approximation to the statistical structure of English. If intentional *S*s respond to sequential dependencies more consistently than incidental *S*s, the difference between the two kinds of learners should increase with the order of approximation. In Exp. II recall was measured for selected items which had appeared in the context of a continuous passage of text. It was expected that frequency of recall would be related, under both conditions of learning, to the degree to which the critical items were determined by the cumulative verbal context of the passage.

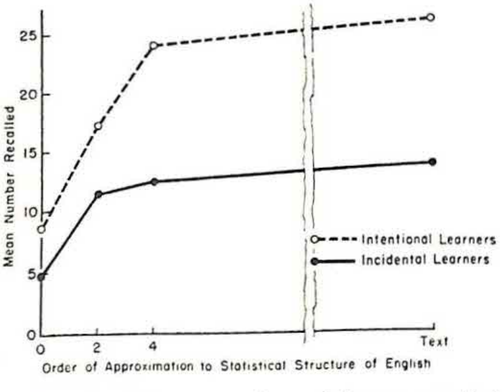


FIG. 1. Mean number of items recalled as a function of instructions and order of approximation.

EXPERIMENT I

Method

Materials. The learning materials consisted of 50-word lists prepared by Miller and Selfridge (1950) in a study of the effects of verbal context on recall. The lists used in the present experiment represented four different orders of approximation to the statistical structure of English: zero order, second order, fourth order, and continuous text. The zero order represents a random arrangement of words drawn from the Thorndike-Lorge word lists. The higher orders were generated by requiring *S*s to complete word sequences of different lengths and linking together the sequences produced by different *S*s. For the second order, the choice of each word was determined by only one preceding word; for the fourth order, each word was obtained in the context of three preceding words. The textual passages were taken directly from current books.

At each order there were two different lists which were used equally often.[2] The lists were recorded, with the words read one by one in a near monotone at a 2-sec. rate. Thus, the sequential links between successive items were not emphasized in the presentation and had to be discovered by *S*.

Procedure. Each list was learned by two different groups of *S*s, one intentional and one incidental. Incidental *S*s were given the assignment to rate the speaker's voice on a number of characteristics (calmness, authoritativeness, naturalness, and cheerfulness). The *S*s were instructed to listen to the entire list and to be prepared to make their ratings on a 5-point scale after the end of the recording. Intentional *S*s were given the same assignment and were, in addition, instructed to memorize as many words as possible, without regard to the order of presentation.

A 5-min. test of free recall was given 30 sec. after the presentation of the list. The *S*s were instructed to write down all the words they could remember, regardless of the order in which the items had appeared in the list. Since there were some recurrent words in each of the higher-order lists, all *S*s were instructed to reproduce each word as often as they thought it had occurred in the list. The ratings of the speaker's voice were made after the end of the test of recall.

---

[2] At each order of approximation, one of the lists was that published by Miller and Selfridge (1950, p. 184 f.). The second list, which had also been used in their study, was kindly furnished to us by G. A. Miller.

Subjects. With four orders of approximation and two kinds of instruction, the experimental design comprised eight conditions. There were 20 *S*s in each condition. The *S*s were undergraduate students at the University of California. The experiment was conducted with small groups averaging four in number. The *S*s were assigned to the experimental conditions in rotation in the order of their appearance in the laboratory, with correction for equal *N*s at the end. On the basis of a postexperimental inquiry three incidental *S*s who had anticipated a test of recall were discarded and replaced.

Results

Amount recalled. The number of words correctly reproduced determined *S*'s recall score. Since *S*s had been instructed to reproduce recurrent items in accordance with their frequency in the list, repeated reproductions of such items were given credit up to the number of actual occurrences in the list. Prior to the over-all analysis the two lists used at each order of approximation were compared. At the zero and second orders of approximation the two lists were found to differ significantly in difficulty; at the two higher orders there were only small differences between the lists. There were, however, no inversions between successive orders. The differences between intentional and incidental learners were consistent from list to list at all orders of approximation.

The mean numbers of items recalled by the different groups are shown in Fig. 1. For intentional learners, recall increases as a negatively accelerated function of the order of approximation. These results agree closely with the findings of Miller and Selfridge (1950), with respect to both the general form of the functional relationship and the absolute level of recall. The curve for incidental learners also shows an initial rise but flattens out earlier than that for the intentional learners. The significance of the differences among the experimental groups was tested by an analysis of variance. Since there was significant heterogeneity of variance, the recall scores were subjected to a Freeman-Tukey square root transformation (Mosteller & Bush, 1954, p. 326 f.) prior to analysis. Both of the main effects are highly significant. For Order of Approximation, $F = 43.54$ ($df = 3$ and 152, $P < .001$); for Instructions, $F = 80.08$ ($df = 1$ and 152, $P < .001$). The interaction, Order of Approximation × Instructions, is also significant ($F = 3.12$, $df = 3$ and 152, $.01 < P < .05$), i.e., there is a reliable divergence between the two kinds of learners as a function of the order of approximation.

Amount recalled increases directly with the order of approximation. Progressive increases in the degree of contextual determination are, however, less beneficial under incidental than under intentional conditions. The negative acceleration of the recall function for both kinds of learners indicates that verbatim recall of a series depends primarily on short-range dependencies rather than on the meaning of the entire passage (cf. Miller & Selfridge, 1950, p. 183). A point is reached beyond which increases in contextual constraint yield diminishing returns. This point is reached earlier for incidental than for intentional learners. The length of contextually determined sequences to which *S*s are likely to respond increases with instructions to learn.

Associative clustering at recall. Additional evidence for the differential effects of context under intentional and incidental conditions comes from an analysis of the sequential

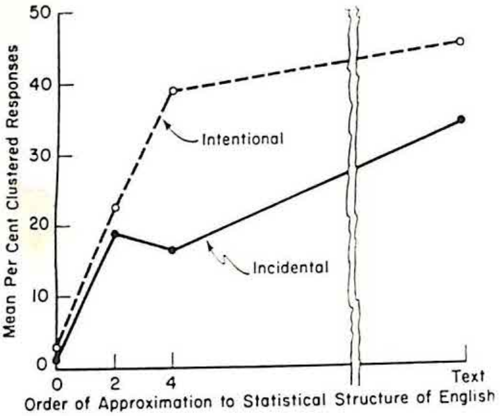


FIG. 2. Mean percentages of clustered responses as a function of instructions and order of approximation.

order in which items were reproduced in recall. Even though the recall instructions permitted *S*s to disregard order, there were many cases in which groups of words were reproduced exactly in the same sequence in which they had appeared in the list. To measure the degree of sequential clustering in recall, the number of items written down in correct sequences of two or more was determined for each *S* and converted into a percentage of his total recall score. Thus, items were scored as clustered if they were adjacent *and* were reproduced sequentially by *S*. Clusters involving recurrent words were included, but a given word was, of course, assigned to only one cluster. A clustered sequence was considered terminated by the intrusion of an error. The mean percentages of clustered responses are shown in Fig. 2. For intentional learners the percentage of clustered responses increases steadily with the order of approximation. Under incidental conditions the increases are neither as extensive nor as regular. Intentional learners give a larger percentage of clustered responses in the recall of all the higher-order lists. This difference is especially pronounced at the fourth order of approximation which represents a high degree of contextual determination and yet falls short of conformity to the conventional rules of sentence construction.

The question must be raised as to the extent to which the results shown in Fig. 2 merely reflect increasing probabilities of chance clusters correlated with rises in the total recall scores. Such an explanation appeared unlikely since the number of possible permutations grows at an extremely fast rate with increases in the total pool of items. Nevertheless, it was decided to apply an empirical correction for chance to the observed measures of clustering. Given a recall score of n items, n numbers between 1 and 50 were chosen in order from a table of random numbers. The percentage of clustered "responses" occurring in the random sequence was then determined according to the same rules as were applied to the actual recall protocols. This procedure was repeated for each of the 160 *S*s in the experiment. The percentages obtained by chance were uniformly low and showed little variation from condition to condition. For the intentional groups, the mean percentages at successively higher orders of approximation were 2.8, 1.8, 2.3, and 0.8. For the incidental groups, the corresponding values were 0.0, 1.0, 1.1, and 1.1. Whenever chance runs did occur, the appropriate percentage was subtracted from *S*'s observed percentage of clustered responses.

Following arc-sine transformation, the corrected percentage scores were subjected to an analysis of variance. Since the distributions at the lowest order of approximation were extremely skewed, heterogeneity of variance remained after the transformation.

The obtained levels of significance are quite high, however, and satisfy the stringent criteria that must be met in such a case (Lindquist, 1953, p. 86). For Order of Approximation, $F = 24.10$ ($df = 3$ and 152, $P < .001$), and for Instructions, $F = 9.48$ ($df = 1$ and 152, $.001 < P < .005$). The interaction of these variables is also significant ($F = 6.05$, $df = 3$ and 152, $P < .01$). When the groups learning the zero-order lists are omitted from the analysis, significant heterogeneity of variance is eliminated. In the reduced analysis both of the main effects remain significant beyond the .01 level. The F ratios are 11.56 ($df = 1$ and 114) and 5.20 ($df = 2$ and 114) for Instructions and Order of Approximation, respectively. The interaction, however, now fails to reach significance ($F = 1.65$). The significant interaction obtained in the first analysis must, therefore, be interpreted with caution.

In a study using lists of different orders of approximation, Deese and Kaufman (1957) found that the correlation between order of recall and position in the original list increased as the characteristics of textual English were approached. The trends in the percentages of clustered responses are consistent with these findings. The *S*s in the present experiment were explicitly instructed to disregard serial order, and the cluster scores are independent of the particular point in the list at which *S* began his reproductions. Thus, the scores reflect correspondences between serial order in learning and order of response wherever they occurred in the course of recall. Such correspondences are present in the records of both kinds of learners, but they reflect the sequential dependencies in the list more clearly under intentional than incidental conditions.

Errors at recall. Table 1 shows the mean numbers of incorrect responses given by the various groups. Under both conditions of learning errors are considerably less frequent at the zero order than at the higher orders of approximation. The rise in errors as a function of the order of approximation is more gradual and regular for the incidental than for the intentional groups. Analysis of variance (after a Freeman-Tukey square root transformation to remove heterogeneity of variance) shows that the number of errors changes significantly with the order of approximation ($F = 6.56$, $df = 3$ and 152, $P < .01$). The difference between intentional and incidental learners is not significant nor is the interaction of order of approximation with instructions ($F < 1$ in both cases).

Incorrect responses as well as correct reproductions tend to increase with the order of approximation. Many of the errors could be identified as synonyms or associates of items in the list and may be attributed to interferences from *S*s' language habits. Presentation of the items in familiar verbal contexts (*a*) permits the transfer of well established verbal habits to the recall of the experimental list, and (*b*) favors the intrusion of incorrect responses associated with these contexts through linguistic

TABLE 1

MEAN NUMBER OF ERRORS AT RECALL AS A FUNCTION OF INSTRUCTIONS AND ORDER OF APPROXIMATION

Order of Approximation	Intentional		Incidental	
	Mean	*SD*	Mean	*SD*
Zero	2.30	2.00	2.75	2.77
Second	6.10	6.17	4.05	4.20
Fourth	5.05	3.23	4.80	6.00
Text	6.65	4.91	6.30	3.64

usage. Hence both amount of recall and number of errors are positively related to order of approximation. At all orders of approximation but the lowest, intentional learners make more errors than the incidental learners although the difference is not significant. The trend is consistent, however, with the findings of earlier studies that instructions to learn enhance the probability of both correct and incorrect intraserial associations (Postman & Adams, 1957).

EXPERIMENT II

Method

Procedure. In the first part of the experiment the "cloze" procedure, developed by Taylor (1953), was used. In this procedure *S*s are presented with a textual passage from which a number of words have been deleted and are required to fill in the missing words. In the present study, a 214-word passage from Julian Huxley's *Man Stands Alone* was used in the cloze procedure. This passage is reproduced in Flesch (1951, p. 16)[3] and was one of those used by Taylor in the development of the cloze technique. Thirty-five words were deleted from the passage. The average separation between successive blanks was 5.0 words, with an *SD* of 1.34. Proper names, articles, pronouns, prepositions, and conjunctions were not deleted. The passage was reproduced on 8.5 × 11-in. sheets and presented to *S*s in 13 separate units. Each unit consisted of a sentence or independent clause, which served to facilitate contextual determination of *S*s' responses. Thus, there was some variation from unit to unit with respect to both the number of words and the number of blanks. The durations of exposure for successive units were designed to allow *S*s to move from blank to blank without delay or interruption. On the basis of preliminary tests the following schedule of exposures was developed: 11 units were exposed for .75 min. each, one unit for 1.25 min., and one unit for 1.75 min. During the exposures, the *S*s used a cardboard strip with a window cutout to cover all the units except the one on which they were working. The *S*s shifted the exposure window from one unit to the next at a signal from *E*. The instructions urged *S*s to fill in all the blanks even if they were not confident about some of their responses.

The cloze procedure was administered to an intentional and an incidental group. The intentional *S*s were instructed to try to remember the words which they used to fill in the blanks, without regard to the order of occurrence in the passage. No learning instructions were given to the incidental *S*s. Thirty seconds after the end of the cloze procedure, a 5-min. test of free recall was given. The *S*s were instructed to write down the words which they had used to fill in the blanks, in any order that occurred to them.

Subjects. There were 50 intentional and 50 incidental *S*s. The procedure was administered to small groups of *S*s averaging three in number. The *S*s were assigned to the two conditions in rotation in the order of their appearance in the laboratory, with correction for equal *N*s at the end. Fifteen intentional and 12 incidental *S*s were discarded and replaced because of failure to fill in all the blanks. One additional incidental *S* was discarded and replaced because of failure to follow instructions.

Results

Cloze scores. The number of blanks filled in correctly defines *S*'s cloze score. The mean cloze score of the Intentional Group was 8.32, with an *SD* of 2.46; for the Incidental Group the mean score was 8.08, with an *SD* of 1.74. The difference between the means does not approach significance ($t = .54$). The distributions of correct completions over blanks were virtually identical under the two conditions. The rank order correlation between the numbers of correct completions is .96.

Contextual determination of responses. The number of correct completions varied widely from blank to blank. These variations reflect differences in the degree of contextual constraint upon *S*s' responses. The number of exactly correct completions does not, however, provide a maximally sensitive index of the degree of contextual determination of responses to any given blank. Many of the incorrect responses were synonyms of the words used in the original passage or could be considered adequate alternatives to the expressions in the text. For this reason the extent of agreement among *S*s, independently of verbatim correspondence with the text, was used as a measure of the degree of contextual determination of the responses to a given blank. For each blank the degree of commonality of responses was determined by comparing each *S*'s response with that of all other *S*s and expressing the frequency of agreements as a proportion of the total number of comparisons. Thus,

---

[3] The last sentence (19 words) in the 233-word passage reproduced in Flesch's book was omitted.

$$C = \frac{A}{N(N-1)/2}$$

where C is the index of commonality, A is the number of agreements, and N is the number of *S*s. The index can vary between 0 and 1. The distribution of the values of C was positively skewed. For the Intentional Group the median C is .19, with a Q of .15; for the Incidental Group the median is .18 and Q is again .15. The distributions of commonality scores obtained under the two conditions are quite similar and there is close correspondence from blank to blank. The rank order correlation between the values of C is .86.

Frequency of usage of completion responses. The words supplied by intentional and incidental *S*s were also quite similar with respect to their frequencies of usage. The Lorge Magazine Count (Thorndike & Lorge, 1944) was used to estimate the frequency of usage of each completion response. The mean log frequency of *S*s' responses was then determined for each blank. For the Intentional Group, the grand mean is 2.81, and the *SD* of the blank means is .48. For the Incidental Group, the corresponding values are 2.80 and .46. As the grand means indicate, words of high frequency of usage were favored by both groups. Again, there is close correspondence between the groups from blank to blank. The rank order correlation between the mean log frequencies of usage by blank is .94. In general, the specific words used by intentional and incidental *S*s to fill in each blank overlapped closely. Even when there was some variation in the number of times each word was used, the values of C and the mean log frequencies of usage remained quite similar.

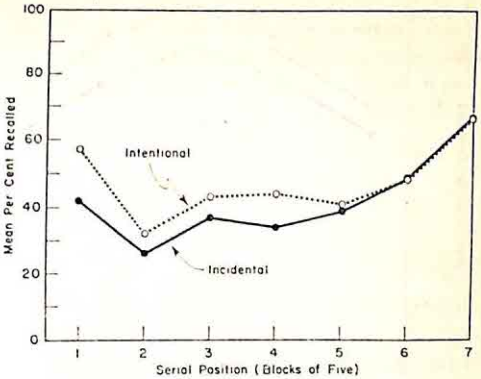


FIG. 3. Percentages of *S*s recalling their responses as a function of serial position, averaged for successive blocks of five.

Amount recalled. The mean number of responses recalled by the Intentional Group was 16.80, with an *SD* of 3.77. For the Incidental Group the mean recall score was 15.32, with an *SD* of 3.71. The difference between the two groups is significant ($t = 1.97$, $df = 98$, $.02 < P < .05$).

Figure 3 shows amount recalled as a function of the serial position of the blanks, averaged for successive blocks of five. For purposes of this plot, responses which were given by the same *S* more than once were omitted. If such a response is re-

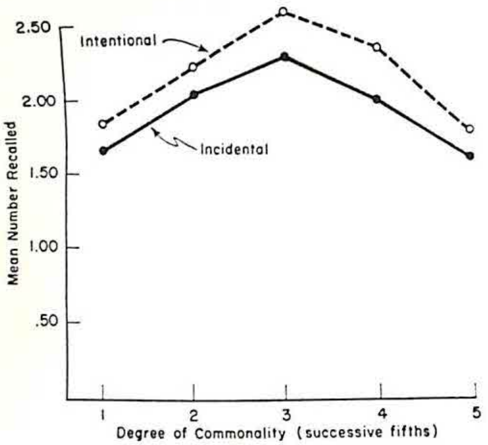


FIG. 4. Mean recall scores as a function of instructions and degree of commonality of responses.

produced only once, its serial position is indeterminate. Such cases occurred with approximately equal frequency under intentional and incidental conditions—the percentage of duplicate responses was 6.5 for the Intentional Group and 7.0 for the Incidental Group. Thus, the omission of duplicate responses does not influence the serial position trends. To adjust for minor variations in frequency of recurrent responses, the figure shows percentages rather than absolute numbers of recalls. In agreement with the results of earlier studies (Postman & Phillips, 1954; Postman & Adams, 1957), the curve for incidental learners shows less of a primacy effect and a relatively greater recency effect than that for intentional learners. The advantage of the intentional group is, therefore, limited to the beginning and middle of the series. The analysis of Deese and Kaufman (1957) suggests that primacy effects in free recall reflect the presence of sequential associations. The difference between the serial position curves of intentional and incidental learners indicates that instructions to learn favored the establishment of sequential associations between responses to successive blanks.

Recall as a function of contextual determination. For purposes of assessing the relation between contextual determination and recall, the blanks were rank ordered with respect to the commonality of *S*s' responses. The rank order was based on the means of the indices of commonality for the intentional and the incidental group. In order to minimize differential effects of serial position, Blanks 1–5 and 31–35 were omitted from this analysis. As Fig. 3 shows, there are moderate rises in the serial position curves between Positions 6 and 30. Within these limits, however, there was no relationship between serial position and degree of commonality (rho = .11). There was also no relationship between degree of commonality and the mean log frequency of usage of *S*s' responses (rho = .11). The 25 blanks included in the analysis were divided into five blocks of five each, in ascending order of degree of commonality. The mean number of responses recalled from each of these blocks is shown in Fig. 4. As the degree of contextual determination increases, amount recalled rises to a peak and then declines at the highest levels of commonality. The trends for the intentional and the incidental group are quite similar. Analysis of variance shows that the differences as a function of degree of commonality are significant ($F = 8.66$, $df = 4$ and 392, $P < .01$). With the early part of the series left out, the over-all difference between the two kinds of learners falls short of significance ($F = 2.77$, $df = 1$ and 98, $.05 < P < .10$). The interaction of instructions with degree of commonality is also not significant ($F < 1$), i.e., variations in contextual constraint have parallel effects on recall under the two conditions.

The results shown in Fig. 4 are only in partial agreement with the original prediction. The initial rises in recall as a function of commonality support the expectation that increases in the degree of contextual determination favor retention. The decline at the highest levels of commonality was not predicted but becomes understandable when we consider the conditions which are responsible for variations in the degree of commonality. Two sources of constraint on *S*s' responses may be distinguished: (*a*) the general context of the passage, and (*b*) the local context. The general context is provided by the content of the passage to which *S*s have been exposed prior to a given blank. Local context refers to the constraints upon *S*s' responses which stem from the specific words or phrases immediately preceding and following the blank. The relative weight of these two sources of constraint varies from blank to blank. When well established language habits are involved, local context alone can give rise to a high degree of commonality of responses. In fact, a high degree of agreement on one specific word is likely to be more heavily determined by the local than the general context. For example, three of the five blanks yielding the highest commonalities of response required the completion of a conventional phrase which is independent of the general context of the passage: holds *true;* having *been;* first *time.* In these examples the italicized word is the dominant completion response which was used by *S*s with virtual unanimity. At the time of free recall, the general context of the passage may be assumed to persist, but specific local contexts are not reinstated. Hence the probability of recall will be directly related to the degree to which the original response was determined by the general context. Recall of responses determined primarily by the local context is likely to be low. The mean frequency of recall for the three items mentioned above was 12.3 under intentional conditions and 6.7 under incidental conditions, whereas the mean frequencies for all blanks were 24.0 and 21.7, respectively. The following interpretation of the trends shown in Fig. 4 is, therefore, suggested. Moderate increases in the commonality of response are attributed primarily to the effects of general context and are accompanied by increases in recall. The highest levels of commonality are likely to reflect heavy determination of responses by local context which is unfavorable to recall. These relationships obtain under both intentional and incidental conditions.

The significance attributed to the general context of the passage may appear to be inconsistent with the finding in Exp. I that recall for the textual passage was little better than for the fourth-order list. The difference between the conditions under which the learning materials were presented in the two experiments may be responsible for this apparent discrepancy. In Exp. I the words in the list were read in a near monotone and presented one by one at a 2-sec. rate. Thus, syntactical units were not differentiated and had to be discovered by *S* in the course of a continuous presentation of individual words. These conditions are not favorable to the cumulative establishment of a general context, and it is likely that only medium-range dependencies were effective during the presentation of the textual passage. In Exp. II, on the other hand, *S*s were presented with complete sentences and clauses and were required to fill in the deletions in a manner consistent with the continuity of the narrative. It is reasonable to assume that the latter conditions are highly favorable to the establishment of a general context. This assumption is confirmed by the results of a postexperimental inquiry in Exp. II in which *S*s were required to state whether they used their knowledge of the general content of the paragraph to help them recall the individual words. Ninety-two per cent of the intentional *S*s and 90% of the incidental *S*s replied in the affirmative. The fact that intentional and incidental *S*s appear to have been equally sensitive to the general context helps to account for the relatively small difference in the amount recalled by the two kinds of learners. When the materials evoke strong differential responses even in the absence of learning instructions the difference between intentional and incidental *S*s is likely to be small (Postman, Adams, & Phillips, 1955).

Recall as a function of frequency of usage. The fact that parts of conventional phrases are recalled infrequently suggests that retention should be poor for any item that is likely to appear in a variety of contexts. Words that are used repeatedly in different contexts usually lack the distinctive connotations that favor strong association with any one context. This interpretation can be checked by an examination of the relationship between the probability of recall and the frequency of usage of *S*s' responses. The higher the frequency of usage of a word the more numerous and varied are the contexts in which it is likely to occur. The present analysis would, therefore, lead us to expect an inverse relationship between the mean frequency of usage of *S*s' responses and the number of recalls.

For Positions 6 through 30 the mean frequency of usage of *S*s' responses was largely independent not only of the degree of commonality but also of serial position (rho = $-.18$ and $-.14$ for the intentional and incidental group, respectively). A rank order correlation was used, therefore, to assess the relationship between mean frequency of usage and recall. For the Intentional Group, the correlation is $-.44$ ($.02 < P < .05$); for the Incidental Group the correlation is $-.51$ ($P < .01$). Under both conditions of learning recall decreases as a function of the frequency of usage of the responses.

Errors at recall. The mean number of incorrect responses given by the Intentional Group on the test of recall was 1.78, with an *SD* of 1.52. For the Incidental Group the mean number was 1.36, with an *SD* of 1.32. The difference between the two groups is not significant ($t = 1.52$). A considerable proportion of the errors—43.3% for the intentional *S*s and 41.2% for the incidental *S*s—were words from the textual passage. Many of the remaining errors appeared to be synonyms or associates of *S*s' responses to the blanks. As in Exp. I, instructions to learn increase the amount recalled without reducing the number of errors. Whatever difference there is points to a positive relationship between amount of recall and the number of overt errors. Thus, instructions to learn again appear to favor the formation of both correct and incorrect associations.

DISCUSSION

The experiments have shown that degree of contextual determination significantly influences recall under both intentional and incidental conditions. Whether or not variations in verbal context interact with the conditions of learning depends on the operations used to measure the effects of contextual determination.

In Exp. I the learning materials consisted of lists which differed widely in the average degree of dependent probability between successive items. The *S*s' task was to reproduce a list of items that had been presented serially. The intentional *S*s' greater sensitivity to verbal context was reflected not only in the amount recalled but also in the degree of associative clustering at recall. While the incidental *S*s also discriminated and reproduced contextually determined sequences, they did so neither as consistently nor as systematically as the intentional learners. Thus, when degree of verbal context is defined as the average dependent probability betweeen the members of a series, it interacts significantly with the conditions of learning. These differential effects of context agree with the results of earlier studies showing that intentional learning is more favorable to the formation of intraserial associations than is incidental learning. When the order of approximation to the statistical structure of English is varied systematically, this difference between the two kinds of learners is brought into clear relief.

Experiment II was concerned with the selective recall of individual responses which had occurred in the context of a continuous passage. The conditions of Exp. II differed from those of Exp. I in two important respects: (*a*) the items which *S*s were required to recall had not formed a series but were scattered throughout the passage; (*b*) *S*s recalled their own responses to the blanks in the passage rather than items prescribed by *E*. In Exp. II degree of context refers, therefore, to variations in dependent probability for selected responses within a list rather than to average differences between lists. Under these conditions frequency of recall depends on the strength of association between specific items and the general context established by the content of the passage. The general context persists at the time of recall and serves to mediate the reproduction of individual items. Responses determined primarily by the local context and parts of conventional phrases unrelated to the general context have a low probability of being recalled. Selective recall of items associated with the general context occurs under both intentional and incidental conditions. The nature of the orienting task insured the establishment of the general context for both groups of *S*s. Hence, variations in the degree of contextual determination produced parallel effects on the probability of recall under intentional and incidental conditions.

SUMMARY

Two studies were reported in which the effects of verbal context on intentional and incidental learning were investigated. In Exp. I intentional and incidental learners were exposed to lists of words representing four orders of approximation to the statistical structure of English. Under both conditions amount retained increased as a negatively accelerated function of the order of approximation to English. The rate of increase was, however, considerably greater for intentional than for incidental learners. The sequence of items in recall reflected contextual dependencies more closely under intentional than under incidental conditions.

In Exp. II recall was tested for selected responses which had been given in the context of a continuous passage of text. Intentional and incidental *S*s filled in the blanks in a passage from which approximately every fifth word had been deleted. Intentional learners recalled more of their own responses than did incidental learners. The degree of agreement among *S*s was used to estimate the amount of contextual constraint on the responses to different blanks. For both groups, the relationship between contextual constraint and recall was curvilinear. Recall was low when contextual constraint was minimal or when the response was part of a conventional phrase. Moderate increases in the commonality of response, which were attributed to the effects of the general context of the passage, were accompanied by increases in the probability of recall.

When verbal context refers to the average dependent probability among the members of a series, increases in the degree of contextual determination are more beneficial under intentional than under incidental conditions. Selective recall of individual items strongly associated with a continuous verbal context occurs under both conditions of learning.

REFERENCES

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

FLESCH, R. F. *How to test readability.* New York: Harper, 1951.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–187.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: VI. Intraserial interference. *J. exp. Psychol.*, 1957, **54**, 153–167.

POSTMAN, L., ADAMS, P. A., & BOHM, A. M. Studies in incidental learning: V. Recall for order and associative clustering. *J. exp. Psychol.*, 1956, **51**, 334–342.

POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and of the method of testing. *J. exp. Psychol.*, 1955, **49**, 1–10.

POSTMAN, L., & PHILLIPS, L. W. Studies in incidental learning: I. The effects of crowding and isolation. *J. exp. Psychol.*, 1954, **48**, 48–56.

TAYLOR, W. L. "Cloze procedure": A new tool for measuring readability. *Journalism Quart.*, 1953, **30**, 415–433.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bur. of Publ., Teacher's Coll., Columbia Univer., 1944.

(Received March 19, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

THE MAGNITUDE OF THE PULFRICH STEREOPHENOMENON AS A FUNCTION OF TARGET VELOCITY[1]

ALFRED LIT[2]

Vision Research Laboratories, University of Michigan

The present experiment deals with a stereoscopic effect that arises whenever a transversely moving object is viewed under conditions of unequal binocular retinal illuminance. The depth effect was first described and analyzed by Pulfrich (1922) and now bears his name. The stereophenomenon can be simply demonstrated by means of a pendulum-bob that is made to oscillate in a frontoparallel plane and on a level with *O*'s eyes. A small target for binocular fixation is positioned in his vertical median plane, directly below the oscillating bob and midway between the end-points of its swing. If a neutral or colored filter is placed in front of one of the eyes while the pendulum-bob is in motion, the bob will appear to rotate out of its plane of oscillation in a horizontal elliptical path that locates the bob nearer than it really is for one direction of stroke, and farther than it really is for the return stroke. The oscillating bob appears to rotate in a clockwise direction (as viewed from above) when the filter is placed before the left eye, and counterclockwise when the filter is placed before the right eye. The stereoeffect becomes noticeable at some threshold difference in binocular retinal illuminance and progressively increases in magnitude as the difference in binocular retinal illuminance is increased.

The explanation of the stereophenomenon given by Pulfrich (1922) is based on a suggestion by Fertsch that increasing differences in binocular retinal illuminance produce increasing differences in the hypothesized visual latent periods of the two eyes. For either eye, the magnitude of the visual latent period (that is, the magnitude of the hypothesized time delay between the onset of stimulation of any given retinal point in the eye and the arousal of the visual effect that signals the position of the stimulus target in space) is assumed to be a reciprocal function of the prevailing level of retinal illuminance. Hence, for any specified position of the moving target in the frontoparallel plane, the delay in the signal from the given stimulated retinal point in the covered eye will be slightly greater than the signal delay from the simultaneously stimulated corresponding retinal point in the uncovered eye. It must be further assumed that the binocular spatial localization of the moving target, at any given moment, is determined by the given pair of retinal points in the two eyes that yields simultaneously aroused binocular signals. Accordingly, to yield simultaneously aroused binocular signals from corresponding retinal points, the onset of stimulation for the eye that is covered by the filter must occur when the moving target is at a position farther behind in its path than the

---

[1] This work was conducted by Project MICHIGAN under Department of the Army Contract (DA-36-039 SC 52654), administered by the United States Army Signal Corps.

[2] Now with Bendix Systems Division, Ann Arbor, Michigan.

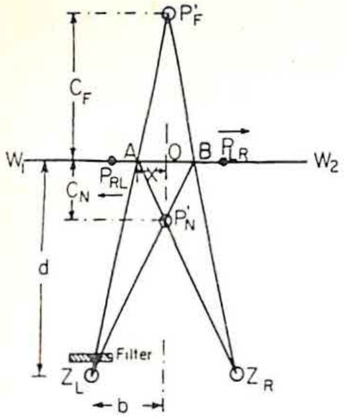


FIG. 1. Geometrical representation of the Pulfrich stereophenomenon indicating the stereoscopic space-image in the horizontal plane of fixation.

position at which the onset of stimulation occurs for the uncovered eye. Thus, simultaneously aroused binocular signals for the transversely moving target are provided by successive pairs of noncorresponding retinal points in the two eyes, and the magnitude of the stereoscopic effect theoretically depends on the amount of the retinal disparity produced as a consequence of the difference in the visual latent periods of the two eyes.

An understanding of the geometric relations involved in the stereophenomenon may be obtained with the aid of Fig. 1. The geometric analysis considers the case of a target that is oscillating in a frontoparallel plane with constant linear velocity, V. The oscillating target is placed at eye level in a frontal plane located at a distance d from the midpoint of the line Z_LZ_R which joins the centers of rotation of the two eyes. The linear path of the target in its plane of oscillation is denoted by W_1W_2. As indicated, the filter is placed in front of the left eye.

Figure 1 is meant to represent the stereoscopic space-image in the horizontal plane of fixation for the case in which the oscillating target, P_{LR}, is moving from left to right, and for the case in which the oscillating target, now designated by P_{RL}, is moving from right to left. The points P'_N and P'_F represent the respective near and far positions at which the oscillating target is supposedly localized by O by use of a binocular fixation target movable in the vertical median plane. The distance OP'_N designates the magnitude of the near displacement, C_N. The distance OP'_F designates the magnitude of the far displacement, C_F.

In accordance with the laws of binocular space discrimination, lines of sight from each eye are drawn through the two respective points of target localization, P'_N and P'_F. The two points of target localization are assumed to have relative positions of depth in the median plane such that the lines of sight drawn through Point P'_N and the lines of sight drawn through Point P'_F intersect the line W_1W_2 at the same two points, A and B. The intersection points A and B theoretically mark the respective positions in the path of the oscillating target at which onset of stimulation occurred in each of the two eyes. Thus, when the target is moving from left to right and appears to be located at the far position, P'_F, the onset of stimulation for the right eye occurred when the target (now at P_{LR}) was located at Point B, and the onset of stimulation for the left eye covered by the filter occurred when the target was located at Point A, a bit farther behind in its path. The time taken for the target to move from B to P_{LR} represents the magnitude of the visual latent period of the right eye, and the time taken for the target to move from A to P_{LR} represents the slightly larger magnitude of the visual latent period of the left eye. Consequently, the time taken for the target to move from A to B represents the difference, Δt_F, in the visual latent periods of the two eyes, based on the far position of target localization, P'_F. Since the velocity of the oscillating target is identical for either direction of movement, it follows by the same reasoning that when the target is moving from right to left and appears to be located at the near position, P'_N, the onset of stimulation for the right eye occurred when the target (now at P_{RL}) was located at Point A, and the onset of stimulation for the left eye covered by the filter occurred when the target was located at Point B, a bit further behind in its path. In this case, the time taken for the target to move from B to A represents the difference, Δt_N, in the visual latent periods of the two eyes, based on the near position of target localization, P'_N. It should be noted that the latency differences based on the near and the far positions of target localization are theoretically equal in magnitude, that is, $\Delta t_N = \Delta t_F$.

If, for any given difference in binocular retinal illuminance, the magnitude of the near and far displacements, C_N and C_F, are experimentally determined, it is possible to calculate the magnitude of the corresponding near and far latency differences, Δt_N and Δt_F, if the linear velocity, V, of the oscillating target is known.

It can be readily seen from similar triangles in Fig. 1 that for target localizations at P'_N and P'_F, respectively,

$$X = bC_N/(d - C_N)$$

and

$$X = bC_F/(d + C_F) \qquad [1]$$

where $X = \frac{1}{2}$ the distance from A to B and $b = \frac{1}{2}$ the distance between the centers of rotation of the two eyes. The distance of the plane of oscillation, d, and the near and far displacements, C_N and C_F, have been previously defined.

For an oscillating target moving with constant linear velocity, V, the time taken for the target to pass through the distance X is given by the formula,

$$t = \frac{X}{V} \qquad [2]$$

The time taken for the target to move from A to B (or B to A) represents the latency difference, Δt, between the two eyes. Since $\Delta t = 2t$, we obtain from Equation 2,

$$\Delta t = \frac{2X}{V} \qquad [3]$$

Substituting for X the respective expressions given in Equation 1, we finally obtain the following relationship between the experimentally determined near and far displacements (C_N and C_F) and the corresponding computed near and far latency differences (Δt_N and Δt_F):

$$\Delta t_N = \frac{2b}{V} \cdot \frac{C_N}{d - C_N}$$

and

$$\Delta t_F = \frac{2b}{V} \cdot \frac{C_F}{d + C_F} \qquad [4]$$

It should be noted from purely geometric considerations that, for any constant difference in binocular retinal illuminance, the magnitude of the stereoscopic effect as measured by C_N and C_F should progressively increase as the linear velocity of the oscillating target is increased, but the corresponding calculated values of Δt_N and Δt_F should remain constant for all target velocities used.

The present experiment is a continuation of a research program (Lit, 1949; Lit & Hyman, 1951) designed to obtain systematic data on some of the important stimulus variables that influence the magnitude of the Pulfrich effect. The aim of the research program is to provide appropriate data that ultimately can be directly related to theories and data concerned with several basic visual functions: (*a*) binocular space discrimination, (*b*) the relationship between the magnitude of the monocular visual latent period and level of retinal illuminance, (*c*) intensity discrimination, (*d*) retinal interaction, and (*e*) color vision. In the present experiment, the effect of target velocity is systematically studied. This is an important variable whose effect on the magnitude of the stereoeffect (that is, on C_N and C_F) can be predicted on the basis of the geometrical theory of the Pulfrich effect. The present experiment thus provides a direct test of the adequacy of the proposed theory. It will also provide data that can be related to classical theory of binocular space discrimination and to theories concerned with retinal interaction effects of moving targets.

APPARATUS AND PROCEDURE

A schematic representation of the apparatus[3] is presented in Fig. 2A. A detailed description is available in a previous report (Lit & Hyman, 1951). The apparatus consists of three major components: (*a*) the oscillating target, (*b*) the fixation target, and (*c*) the lighting and screening units.

The *O* is seated in a dark room (*D*) and binocularly observes the fixation target (*FT*) and the oscillating target (*OT*) through a pair of circular artificial pupils (*E*) that are 2.5 mm. in diameter and adjustable for interpupillary separation. The artificial pu-

---

[3] The apparatus was originally constructed at Pupin Laboratories, Columbia University, partially through funds from a research grant-in-aid generously provided by the American Academy of Optometry.

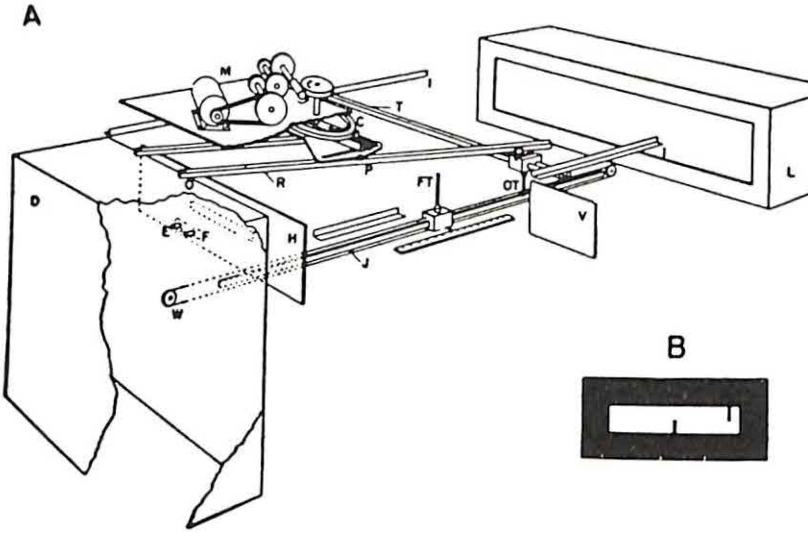


FIG. 2. Schematic representation of the apparatus and O's view of stimulus targets. A. The O is seated in a dark room (D) and binocularly observes the fixation target (FT) located in the lower visual field and the oscillating target (OT) located in the upper visual field through a pair of artificial pupils (E). Movement of the oscillating target in a frontoparallel plane 100 cm. from O's eyes can be varied over a wide range of constant linear velocities. The fixation target in O's vertical median plane can be moved either toward or away from his eyes by means of a pulley wheel (W) located in the dark room. Background illumination is provided by a lightbox (L). The retinal illuminance of each eye is controlled by neutral density filters placed in the pair of filter boxes (F). Horizontal (H) and vertical (V) screens provide a constant rectangular field of view. B. The upper rod is the oscillating target; the lower rod is the fixation target. (From Lit & Hyman, 1951.)

pils are attached to eye-tubes which are mounted on the inner wall of the dark room. In front of each eye-tube, a filter box (F) is mounted on the outer wall so that E can control the retinal illuminance of each eye by combinations of neutral density filters. The O's head is kept immobilized by means of chin and forehead rests.

The oscillating target (OT) is a blackened steel rod $\frac{1}{8}$ in. in diameter. It is vertically suspended downward to eye level from a Jacobs chuck in which it is retained. The chuck is centrally mounted on the undersurface of a supporting carriage which rides on horizontal tracks (T) located in a frontal plane at a distance of 100 cm. from O's eyes. The carriage receives its movement from a horizontally oscillating drive-rod (R) which is pivoted at Position O, a point located in O's median plane directly above the midpoint of the line joining the two eyes. Power for the drive-rod is provided by a cam-regulated mechanism (C) which converts constant angular velocity into reciprocating linear velocity, with the central 90% of stroke at constant speed. The power to the drive-rod is applied at a vertical pivot point (P) permanently mounted on the drive-rod. The electrically driven gear train (M) shown in Fig. 2A was replaced by a Zero-Max (Revco, Inc., Model 143) variable speed transmission device to allow adjustments of the linear velocity of the oscillating target over a wide range of values. Calibration of the transmission device was achieved by measuring the time required for the supporting carriage of the oscillating target to move through a fixed distance of 10 cm. in the central region of the elevated tracks (T) on which it rides. The time measurements were performed with an Electronic Precision Chronoscope (Wichita Apparatus Supply, Inc., Model 251). Thus, the linear velocity of the oscillating target is specified for all positions of the lever arm of the speed control link.

The fixation target (FT) is a blackened steel rod $\frac{1}{8}$ in. in diameter. It is held vertically upright to eye level in a Jacobs chuck that is mounted on the upper surface of a supporting carriage located below eye level. The comparison rod and its supporting carriage are movable along a horizontal metal track (J) located in O's vertical median plane. By means of a pulley wheel (W) located in the dark room, O can adjust the position of the fixation target in a direction either towards or away from his eyes. The distance of the fixation target from O's eyes, as measured along the metal track, can be read by E from a scale calibrated in millimeters. The use of a vernier index permits E to estimate the distance of the fixation target to within 0.01 cm. The height of the upper end of the fixation target is set on a level with O's eyes. Thus, when the oscillating target is at a position directly above the fixation target, the targets appear contiguous in O's vertical median plane. At this distance (100 cm.) from O's eyes, the diameter of each rod subtends a visual angle of 10.9 minutes of arc.

Uniform background illumination is provided by ten 150-w. frosted lamps that are appropriately mounted in an asbestos lined, galvanized iron lightbox (L). The lightbox is located in a frontal plane 250 cm. from O's eyes. Lamp voltage is maintained constant (to within ±1.0%) at 124 v. AC by means of an automatic constant-voltage output regulator. The illuminated surface is a white matte screen that is attached to the inner rear wall of the lightbox. The surface has a luminance of 854 ft.-L. as measured with a Macbeth illuminometer. The color temperature at the given voltage is 2735° K. With the 2.5-mm. artificial pupil in use, the retinal illuminance without filters is 14359 trolands or 4.16 log trolands. The field of view in the vertical direction is kept constant at 4.2° by a horizontal slit (26 cm. × 1.5 cm.) cut at eye level in a black vertical screen (H) located 21 cm. in front of O. The field of view in the horizontal direction is kept constant at 21.6° by means of a pair of vertical screens (V) (only one screen is shown in Fig. 2A) adjusted symmetrically in the plane of the oscillating target, 0.5° beyond the end-points of its reciprocating stroke. The view of the targets as seen by O is shown in Fig. 2B.

Two trained graduate students who were emmetropic served as paid Os. The monocular visual acuity of each O was better than 20/20, and the ductions of each were normal at both distance and near. At a fixation distance of 100 cm., the interpupillary separation for Observer F. C. was 6.20 cm. and that for Observer M. M. was 6.70 cm. At this fixation distance, the phoria for Observer F. C. was 3^Δ exophoria and that for Observer M. M. was 1^Δ esophoria.

Daily practice sessions were held for a period of about a month during which Os were trained in the procedure of localizing the apparent near and far positions of the oscillating target at various target velocities and under different amounts of unequal binocular retinal illuminances, $\log(E_R/E_L)$. In performing his settings, O continuously fixates the upper end of the movable fixation rod and adjusts this rod in the vertical median plane until it appears to lie directly below the near and far paths of the oscillating target. The apparent near and far positions of the oscillating target are each localized first when the fixation rod is moved away, and again when it is moved toward O. In this way, multiple pairs of determinations of C_N and C_F can be obtained under any given set of viewing conditions. With filters of equal optical density in front of the eyes, only a single path of the oscillating target was reported; that is, no Pulfrich effect was elicited at any given velocity under conditions of equal binocular retinal illuminance.

Settings for the apparent near and far positions of the oscillating target were obtained from both Os at each of 11 target velocities, V: 2.59, 5.90, 8.16, 10.28, 13.76, 19.96, 26.86, 35.56, 45.01, 55.53, and 68.17 cm./sec. For target movement in a frontoparallel plane located 100 cm. from O's eyes, these values of linear velocity correspond to the following angular velocities: 1.49, 3.38, 4.68, 5.89, 7.88, 11.44, 15.39, 20.37, 25.78, 31.81, and 39.05 deg./sec. In a given experimental session only one target velocity was used and five pairs of settings (10 readings each for C_N and C_F) were obtained from each O for each of four conditions of increasing inequality of binocular retinal illuminance, $\log(E_R/E_L)$. The retinal illuminance of the left eye ($\log E_L$) was held constant at 2.06 log trolands by use of a neutral filter of optical density 2.10. The retinal illuminance of the right eye ($\log E_R$) was successively increased by use of neutral filters of optical densities 1.98, 1.52, 1.03, and 0.52. Thus, the four values of $\log(E_R/E_L)$ used at each of the 11 target velocities were: 0.12, 0.58, 1.07, and 1.58. A total of 22 experimental sessions was held for each O. A counterbalanced order was introduced for target velocity. That is, for the first 11 sessions the target velocity was presented in order of increasing magnitude, and for the second 11 sessions the target velocity was presented in decreasing order.

RESULTS

The results for both Os are presented in Table 1. Each entry of

TABLE 1

DEPTH DISPLACEMENTS (IN CENTIMETERS) OBTAINED AT EACH OF 11 TARGET VELOCITIES UNDER FOUR CONDITIONS OF UNEQUAL BINOCULAR RETINAL ILLUMINANCE, LOG (E_R/E_L)

Target Velocity Cm. per Sec.)	log $(E_R/E_L) = 0.12$				log $(E_R/E_L) = 0.58$				log $(E_R/E_L) = 1.07$				log $(E_R/E_L) = 1.58$			
	F. C.		M. M.		F. C.		M. M.		F. C.		M. M.		F. C.		M. M.	
	C_N	C_F	C_N	C_F	C_N	C_F	C_N	C_F	C_N	C_F	C_N	C_F	C_N	C_F	C_N	C_F
2.59	0.13	0.20	0.33	0.39	0.48	0.46	0.63	0.59	0.74	0.75	0.85	0.82	0.95	0.98	1.07	1.03
5.90	0.27	0.19	0.44	0.42	0.80	0.70	0.85	0.94	1.28	1.16	1.37	1.35	1.63	1.48	1.93	1.93
8.16	0.41	0.27	0.35	0.62	1.12	0.97	1.09	1.22	1.55	1.48	1.67	1.63	1.99	1.90	2.17	2.27
10.28	0.54	0.21	0.49	0.70	1.43	1.07	1.26	1.50	2.05	1.68	2.04	2.18	2.45	2.14	2.49	2.77
13.76	0.83	0.38	0.61	1.01	1.91	1.36	1.58	1.88	2.57	2.09	2.42	2.77	3.08	2.79	3.24	3.45
19.96	1.31	0.19	0.61	1.29	2.69	1.61	2.15	2.64	3.52	2.73	3.30	3.76	4.12	3.45	3.74	4.99
26.86	2.01	0.82	0.72	1.67	3.54	2.23	2.67	3.60	4.64	3.24	3.74	5.18	5.98	4.54	4.80	6.38
35.56	2.58	0.57	0.87	2.23	4.77	2.21	2.42	4.21	6.07	3.49	4.52	6.74	7.41	5.27	5.77	8.88
45.01	3.15	0.58	0.97	2.01	5.45	2.50	3.43	4.64	7.32	4.33	5.37	7.85	9.08	6.35	6.82	9.48
55.53	3.72	0.05	0.77	2.79	6.62	2.13	4.41	6.63	8.93	3.73	6.88	9.45	10.70	6.21	8.42	12.33
68.17	4.22	0.02	0.60	5.70	7.50	1.45	4.46	9.89	10.66	3.46	7.64	13.30	12.57	5.95	9.40	15.74

Note.—The retinal illuminance of the left eye, log E_L, was kept constant at 2.06 log trolands. C_N and C_F refer, respectively, to the near and far displacements of a black vertical rod oscillating in a frontoparallel plane located 100 cm. from *O*'s eyes. Each entry for the two *O*s (F. C. and M. M.) is based on the mean of 20 settings. The thickness of the stimulus target used was 0.125 in.

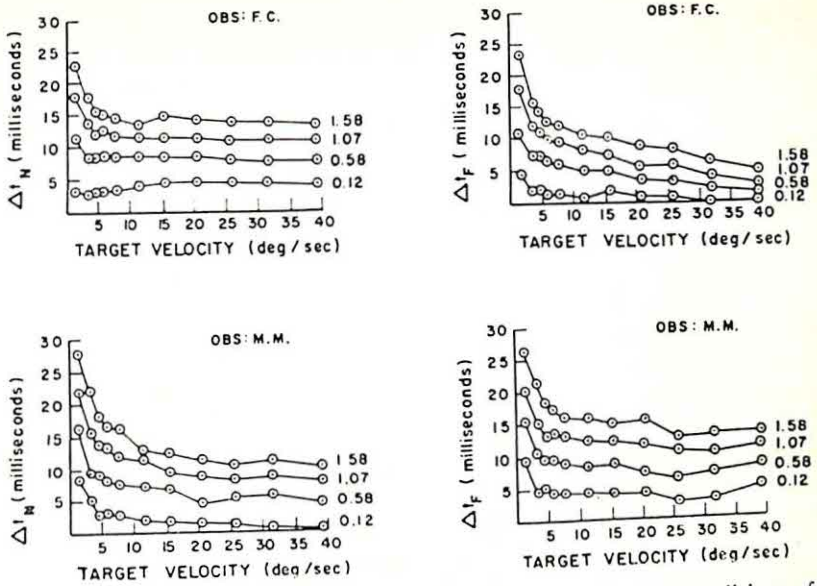


FIG. 3. Latency differences as a function of target velocity for four conditions of unequal binocular retinal illuminance. The latency differences (Δ_N and Δ_F) were computed from Equation 4 for the corresponding near and far displacements (C_N and C_F) given in Table 1. The number accompanying each curve represents the prevailing magnitude of log (E_R/E_L), where the retinal illuminance of the left eye, log E_L, is kept constant at 2.06 log trolands. Each point is based on the mean of 20 readings.

C_N and C_F represents the mean value (in centimeters) of two sets of 10 readings obtained at each target velocity under each of the four specified conditions of unequal binocular retinal illuminance.

Inspection of Table 1 reveals that, for any given value of log (E_R/E_L), C_N and C_F progressively increase as target velocity is increased. C_N and C_F also progressively increase as log (E_R/E_L) is increased at any given target velocity. A characteristic individual difference in performance should be noted: for Observer F. C. the values of C_N are consistently larger than the corresponding value of C_F; for Observer M. M., contrariwise, the values of C_N are consistently smaller than the corresponding values of C_F.

To facilitate analysis of these data in terms of the geometrical theory of the Pulfrich effect, the corresponding latency differences, Δt_N and Δt_F, have been computed from Equation 4 for each set of values of C_N and C_F given in Table 1. The results of the computation are shown graphically in

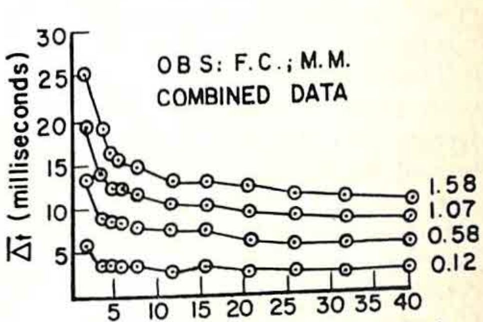


FIG. 4. Average latency difference as a function of target velocity for four conditions of unequal binocular retinal illuminance, log (E_R/E_L). The average latency differences, $\overline{\Delta t}$, were obtained by combining the values of Δt_N and Δt_F in Fig. 3 for both *O*s. Thus, each point is based on the mean of 40 readings.

Fig. 3 where Δt_N and Δt_F (in milliseconds) are plotted for each O as a function of target velocity (in deg. per sec.) with $\log (E_R/E_L)$ serving as parameter. A similar plot of the averages of the near and far latency differences, $\overline{\Delta t}$, for the combined data of both observers is shown in Fig. 4.

The curves in Fig. 3 and 4 for $\log (E_R/E_L) = 0.12$ show that the computed values of Δt_N, Δt_F, and $\overline{\Delta t}$ remain essentially constant as target velocity is progressively increased. A slight upturn occurs, however, at the lowest target velocity for the curves representing this condition of slightly unequal binocular retinal illuminance. For the remaining curves, the upturn above their respective constant levels becomes more marked and occurs at progressively higher values of target velocity as the magnitude of inequality of binocular retinal illuminance is increased; for the curves representing $\log (E_R/E_L) = 1.58$, the computed latency differences seem to be independent of target velocity for velocities greater than about 20 deg./sec. It is also to be noted from Fig. 3 and 4 that at any given target velocity the computed latency differences progressively increase in magnitude as $\log (E_R/E_L)$ is increased.

The individual difference previously noted with respect to the relative magnitudes of C_N and C_F also prevails with respect to the relative magnitudes of Δt_N and Δt_F: for Observer F. C., the values of Δt_N are consistently larger than the corresponding values of Δt_F; for Observer M. M., the values of Δt_N are consistently smaller than the corresponding values of Δt_F.

DISCUSSION

Of all stimulus factors known to influence the magnitude of the Pulfrich stereophenomenon, target velocity has been the variable most frequently studied by previous investigators (Arndt, 1930; Banister, 1932; Engelking & Poos, 1924; Holz, 1934; Liang & Piéron, 1947; Pulfrich, 1922; Wölfflin, 1925). Although in most of these experiments displacement settings were obtained for only limited ranges of target velocity or for only a single condition of unequal binocular retinal illuminance, the results invariably showed that the values of C_N and C_F were smallest at the lowest target velocity and progressively increased as target velocity was increased. The corresponding absolute magnitudes of the computed latency differences, Δt_N and Δt_F, reported by these investigators, revealed considerable variability. When a dense blue filter was placed in front of one eye, Arndt (1930) found a computed latency difference of 1917 msec. for a target velocity of 0.04 deg. per sec. and a computed latency difference of 110 msec. for a target velocity of 6.67 deg. per sec. These data thus show a 17-fold decrease in latency difference as target velocity undergoes a 170-fold increase in magnitude. In contrast, the data of Holz (1934) revealed that, for a given blue filter, a computed latency difference of 97 msec. was obtained for a target velocity of 0.65 deg. per sec. and a computed latency difference of 16 msec. was obtained for a target velocity of 23.23 deg./sec. In this case, latency difference undergoes only about a 6-fold change. Holz also reported that the computed latency difference showed no further decrease in magnitude (below 16 msec.) as target velocity was respectively increased to values of 40.42 and 130.53 deg./sec.

The data of the present experiment (that is, the curves given in Fig. 4) seem to show better agreement with predictions based on the geometrical theory of the Pulfrich effect. For each of the specified differences of binocular retinal illuminance used, the respective magnitude of the computed average latency difference, $\overline{\Delta t}$, remains essentially constant as target velocity is systematically varied. Discrepancies appear primarily at low target velocities, particularly for the experimental curves representing large values of log (E_R/E_L). These discrepancies reflect the fact that as target velocity is progressively decreased the values of C_N and C_F obtained for the given curve become systematically slightly larger than the respective theoretical values required by Equation 4 to yield a given constant $\overline{\Delta t}$ for all target velocities.

A similar discrepancy for low target velocities occurs in the so-called sensation–time (*Empfindungszeit*) experiments of Fröhlich (1923) in which a vertical slit of light is moved in a frontoparallel plane directly behind a horizontal opening in a screen. To *O* seated directly in front of the screen, the moving target typically does not appear to come into view at the entrance edge of the horizontal opening but rather at some small lateral distance within the border of the screen aperture. Fröhlich attributed this effect to the visual latent period (*die Empfindungszeit*). He proposed that the magnitude of the sensation–time for any given set of observation conditions could be computed from the time difference between the actual and seen appearance of the moving target, that is, sensation–time $= d/v$, where d is the lateral distance from the entrance edge at which the vertical target first appears to come into view and v is the linear velocity of the target in its frontoparallel path. Experiments on the effects of target velocity (e.g., Holz, 1934) yielded sensation–time vs. target velocity curves for various target luminances that also characteristically show an upturn at low target velocities. That is, as target velocity is progressively decreased, the respective lateral distances from the entrance edge at which the target is localized by *O* become progressively smaller but at a rate considerably slower than that required to yield a given constant computed sensation–time for all target velocities. Holz (1934) obtained sensation–time measures under stimulus conditions that were identical to those prevailing in his experiment on the Pulfrich effect, that is, target velocities and target luminances were identical in the two studies. His results showed that, for each of the given target velocities, the algebraic difference between the magnitude of the sensation–time for the target having the lower luminance and the magnitude of the sensation–time for the target having the higher luminance yielded a value that was numerically equal to the respective computed latency difference obtained in his experiment on the Pulfrich effect.

Additional experiments on the effects of target velocity are required to clarify the reasons for the upturns in the latency difference vs. target velocity curves of Fig. 4. Of particular interest in this connection would be the results of additional experiments in which the thickness of the oscillating target is systematically varied and in which specified differences of binocular retinal illuminances are produced at many basic levels of illuminance.

The present experiment also provides data on the effects of specified differences in binocular retinal illuminance on the magnitude of the near and far displacements and their corresponding computed latency differences. The data show that, for each of the given target velocities, displacements C_N and C_F and their corresponding calculated latency differences, Δt_N and Δt_F, progressively increase as log (E_R/E_L) is increased. As in the case of similar data obtained in an earlier experiment (Lit, 1949) the effects of variations in the magnitude of log (E_R/E_L) can be accounted for if the assumption is made that the hypothesized absolute visual latent period, t, is an inverse function of retinal illuminance, log E. A schematic representation of this relationship is shown in Fig. 5. In this figure, log E_L represents the constant retinal illuminance of the left eye, and log E_{R1}, log E_{R2}, and log E_{R3} represent the increased retinal illuminance successively produced in the right eye. It can be readily seen with the aid of Fig. 5 that the difference in absolute latency (Δt) theoretically increases as the difference in binocular retinal illuminance [log

Fig. 3 where Δt_N and Δt_F (in milliseconds) are plotted for each O as a function of target velocity (in deg. per sec.) with $\log (E_R/E_L)$ serving as parameter. A similar plot of the averages of the near and far latency differences, $\overline{\Delta t}$, for the combined data of both observers is shown in Fig. 4.

The curves in Fig. 3 and 4 for $\log (E_R/E_L) = 0.12$ show that the computed values of Δt_N, Δt_F, and $\overline{\Delta t}$ remain essentially constant as target velocity is progressively increased. A slight upturn occurs, however, at the lowest target velocity for the curves representing this condition of slightly unequal binocular retinal illuminance. For the remaining curves, the upturn above their respective constant levels becomes more marked and occurs at progressively higher values of target velocity as the magnitude of inequality of binocular retinal illuminance is increased; for the curves representing $\log (E_R/E_L) = 1.58$, the computed latency differences seem to be independent of target velocity for velocities greater than about 20 deg./sec. It is also to be noted from Fig. 3 and 4 that at any given target velocity the computed latency differences progressively increase in magnitude as $\log (E_R/E_L)$ is increased.

The individual difference previously noted with respect to the relative magnitudes of C_N and C_F also prevails with respect to the relative magnitudes of Δt_N and Δt_F: for Observer F. C., the values of Δt_N are consistently larger than the corresponding values of Δt_F; for Observer M. M., the values of Δt_N are consistently smaller than the corresponding values of Δt_F.

DISCUSSION

Of all stimulus factors known to influence the magnitude of the Pulfrich stereophenomenon, target velocity has been the variable most frequently studied by previous investigators (Arndt, 1930; Banister, 1932; Engelking & Poos, 1924; Holz, 1934; Liang & Piéron, 1947; Pulfrich, 1922; Wölfflin, 1925). Although in most of these experiments displacement settings were obtained for only limited ranges of target velocity or for only a single condition of unequal binocular retinal illuminance, the results invariably showed that the values of C_N and C_F were smallest at the lowest target velocity and progressively increased as target velocity was increased. The corresponding absolute magnitudes of the computed latency differences, Δt_N and Δt_F, reported by these investigators, revealed considerable variability. When a dense blue filter was placed in front of one eye, Arndt (1930) found a computed latency difference of 1917 msec. for a target velocity of 0.04 deg. per sec. and a computed latency difference of 110 msec. for a target velocity of 6.67 deg. per sec. These data thus show a 17-fold decrease in latency difference as target velocity undergoes a 170-fold increase in magnitude. In contrast, the data of Holz (1934) revealed that, for a given blue filter, a computed latency difference of 97 msec. was obtained for a target velocity of 0.65 deg. per sec. and a computed latency difference of 16 msec. was obtained for a target velocity of 23.23 deg./sec. In this case, latency difference undergoes only about a 6-fold change. Holz also reported that the computed latency difference showed no further decrease in magnitude (below 16 msec.) as target velocity was respectively increased to values of 40.42 and 130.53 deg./sec.

The data of the present experiment (that is, the curves given in Fig. 4) seem to show better agreement with predictions based on the geometrical theory of the Pulfrich effect. For each of the specified differences of binocular retinal illuminance used, the respective magnitude of the computed average latency difference, $\overline{\Delta t}$, remains essentially constant as target velocity is systematically varied. Discrepancies appear primarily at low target velocities, particularly for the experimental curves representing large values of log (E_R/E_L). These discrepancies reflect the fact that as target velocity is progressively decreased the values of C_N and C_F obtained for the given curve become systematically slightly larger than the respective theoretical values required by Equation 4 to yield a given constant $\overline{\Delta t}$ for all target velocities.

A similar discrepancy for low target velocities occurs in the so-called sensation–time (*Empfindungszeit*) experiments of Fröhlich (1923) in which a vertical slit of light is moved in a frontoparallel plane directly behind a horizontal opening in a screen. To *O* seated directly in front of the screen, the moving target typically does not appear to come into view at the entrance edge of the horizontal opening but rather at some small lateral distance within the border of the screen aperture. Fröhlich attributed this effect to the visual latent period (*die Empfindungszeit*). He proposed that the magnitude of the sensation–time for any given set of observation conditions could be computed from the time difference between the actual and seen appearance of the moving target, that is, sensation–time $= d/v$, where d is the lateral distance from the entrance edge at which the vertical target first appears to come into view and v is the linear velocity of the target in its frontoparallel path. Experiments on the effects of target velocity (e.g., Holz, 1934) yielded sensation–time vs. target velocity curves for various target luminances that also characteristically show an upturn at low target velocities. That is, as target velocity is progressively decreased, the respective lateral distances from the entrance edge at which the target is localized by *O* become progressively smaller but at a rate considerably slower than that required to yield a given constant computed sensation–time for all target velocities. Holz (1934) obtained sensation–time measures under stimulus conditions that were identical to those prevailing in his experiment on the Pulfrich effect, that is, target velocities and target luminances were identical in the two studies. His results showed that, for each of the given target velocities, the algebraic difference between the magnitude of the sensation–time for the target having the lower luminance and the magnitude of the sensation–time for the target having the higher luminance yielded a value that was numerically equal to the respective computed latency difference obtained in his experiment on the Pulfrich effect.

Additional experiments on the effects of target velocity are required to clarify the reasons for the upturns in the latency difference vs. target velocity curves of Fig. 4. Of particular interest in this connection would be the results of additional experiments in which the thickness of the oscillating target is systematically varied and in which specified differences of binocular retinal illuminances are produced at many basic levels of illuminance.

The present experiment also provides data on the effects of specified differences in binocular retinal illuminance on the magnitude of the near and far displacements and their corresponding computed latency differences. The data show that, for each of the given target velocities, displacements C_N and C_F and their corresponding calculated latency differences, Δt_N and Δt_F, progressively increase as log (E_R/E_L) is increased. As in the case of similar data obtained in an earlier experiment (Lit, 1949) the effects of variations in the magnitude of log (E_R/E_L) can be accounted for if the assumption is made that the hypothesized absolute visual latent period, t, is an inverse function of retinal illuminance, $\log E$. A schematic representation of this relationship is shown in Fig. 5. In this figure, $\log E_L$ represents the constant retinal illuminance of the left eye, and $\log E_{R1}$, $\log E_{R2}$, and $\log E_{R3}$ represent the increased retinal illuminance successively produced in the right eye. It can be readily seen with the aid of Fig. 5 that the difference in absolute latency (Δt) theoretically increases as the difference in binocular retinal illuminance [log

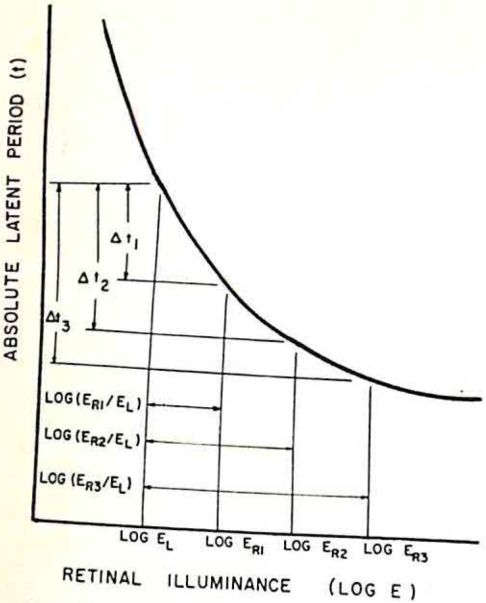


FIG. 5. The hypothesized absolute visual latent period (t) as a function of retinal illuminance ($\log E$). The curve represents an assumed relationship proposed to account for the experimental fact that, for the given constant retinal illuminance of the left eye ($\log E_L$), latency difference (Δt) increases progressively as the difference in binocular retinal illuminance [$\log (E_R/E_L)$] is increased.

(E_R/E_L)] is increased. The specific rate of increase theoretically depends, of course, on the initial magnitude selected for the retinal illuminance of the left eye ($\log E_L$).

When the data of the present experiment are plotted to show how $\overline{\Delta t}$ varies as a function of $\log (E_R/E_L)$, with target velocity serving as parameter, the obtained curves (not given here) are in quantitative agreement with predictions based on the analysis of Fig. 5. The curves do not, however, overlap nor show the same shape for all target velocities. The curves representing low target velocities (below 10 deg./sec.) are considerably displaced progressively upward on the ordinate axis. The curves show a relatively more rapid rise in $\overline{\Delta t}$ as $\log (E_R/E_L)$ is increased, an effect, of course, which reflects the lack of parallelism exhibited in the curves of Fig. 4.

From data obtained in an earlier experiment (Lit, 1949), it was possible to determine an empirical equation which describes the relationship existing between the hypothesized absolute visual latent period (t) and level of retinal illuminance ($\log E$). The necessary computations were deferred pending the outcome of additional experiments concerned with evaluating the effects of systematic variations in distance of observation of the oscillating target, target velocity, and target thickness. It has since been established (Lit & Hyman, 1951) that latency difference is independent of distance of target oscillation. The present experiment demonstrates that, for small differences of binocular retinal illuminance, latency difference is independent of target velocity over a very wide range of velocities. The effects of systematic variations in target thickness will be studied next and the results reported separately.

Finally, mention should be made of the localization error that exists for the moving target when viewed under conditions of equal binocular retinal illuminance. For *Os* used in the previous experiments (Lit, 1949; Lit & Hyman, 1951) the oscillating target appeared to be moving in a frontoparallel plane located nearer than the plane defined by the "true" distance of the oscillating rod. The localization error occurred at all levels of equal retinal illuminance (Lit, 1949) and at all distances of observation (Lit & Hyman, 1951). A similar localization error occurred for *Os* used in the present experiment. At all target velocities, the oscillating target appeared nearer than the true plane of oscillation for Observer F. C. and farther than the true plane of oscillation for Observer M. M. It should be pointed out that the magnitude of the localization error for each *O* was insufficient to establish equality between the corresponding near and far computed latency differences. Thus, when the corresponding values of C_N and C_F were "corrected" for the localization error, it still turned out that $\Delta t_N > \Delta t_F$ for Observer F. C. and that $\Delta t_N < \Delta t_F$

for Observer M. M., particularly for the displacements produced by the larger differences of binocular retinal illuminance. In all cases, the computed average latency difference, $\overline{\Delta t}$, remained virtually unchanged when the respective values of Δt_N and Δt_F were each "corrected" for the localization error.

SUMMARY

When filters of unequal optical density are placed in front of the two eyes, a target which is actually oscillating in a frontoparallel plane appears nearer than it really is for one direction of stroke and farther than it really is for the return stroke (Pulfrich stereophenomenon). Measurements of the near and far displacements of a black vertical rod have been obtained for a wide range of target velocities under each of several conditions of unequal binocular retinal illuminance.

The experimental data show that, for any given difference in binocular retinal illuminance, the near and far displacements progressively increase as target velocity is increased. The data show also that, for any given target velocity, the near and far displacements progressively increase as the difference in binocular retinal illuminance is increased.

The obtained results are analyzed in terms of an hypothesized absolute visual latent period whose magnitude is assumed to be an inverse function of level of retinal illuminance. The results are shown to be in good quantitative agreement with predictions based on the geometrical theory of the Pulfrich effect. Discrepancies at low target velocities are noted and discussed.

REFERENCES

ARNDT, G. Über die Abhängigkeit des Stereoeffektes von der Geschwindigkeit der bewegten Marke. *Z. Biol.*, 1930, **90,** 574–588.

BANISTER, H. Retinal action time. In *Report of a joint discussion on vision,* London: Physical Society, 1932. Pp. 227–235.

ENGELKING, E., & POOS, F. Über die Bedeutung des Stereophänomens für die isochrome und heterochrome Helligkeitsvergleichung. *Arch. f. Ophthal.*, 1924, **114,** 340–379.

FRÖHLICH, F. W. Über die Abhängigkeit der Empfindingszeit und des zeitlichen Verlaufes der Gesichtsempfindung von der Intensität, Dauer und Geschwindigkeit der Belichtung. *Z. Sinnesphysiol.*, 1923, **55,** 1–46.

HOLZ, J. Der Stereoeffekt Pulfrich's und die Empfindungszeit. *Z. Biol.*, 1934, **95,** 502–516.

LIANG, T., & PIÉRON, H. Recherches sur la latence de la sensation lumineuse par la méthode de l'effect chronostéréoscopique. *Année Psychol.*, 1947, **43–44,** 1–53.

LIT, A. The magnitude of the Pulfrich stereophenomenon as a function of binocular differences of intensity at various levels of illumination. *Amer. J. Psychol.*, 1949, **62,** 159–181.

LIT, A., & HYMAN, A. The magnitude of the Pulfrich stereophenomenon as a function of distance of observation. *Amer. J. Optom.*, 1951, **28,** 564–580.

PULFRICH, C. Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. *Naturwissenschaften*, 1922, **10,** 553–564; 569–574; 596–601; 714–722; 735–743; and 751–761.

WÖLFFLIN, E. Untersuchungen über den Pulfrichschen Stereoeffekt. *Arch. Augenheilk.*, 1925, **95,** 167–179.

(Received March 23, 1959)

TABLE 1

Stimulus Materials and Critical Instructions

Cond.	Set	Color	Class	Critical Instructions
I	outer	B (black)	C (consonants)	"letters" or "numbers"
	inner	R (red)	D (digits)	
II	*either*			"letters" or "numbers"
	outer	B	C	
	inner	B	D	
	or			
	outer	B	D	
	inner	B	C	
III	same as Cond. II			"outer" or "inner"
IV	*either*			"red" or "black"
	outer	R	C	
	inner	B	C	
	or			
	outer	B	C	
	inner	R	C	

shutter speed .1 sec.). The room was dimly lit and *S* recorded his report immediately after each presentation on a paper slip with a vertical line down the center, which *E* then removed. The tachistoscope was operated by a solenoid controlled by a pulse recorded on the bottom track of a length of magnetic tape. The top track of this control tape carried the two instructions (see below). The tape was fed through a two-channel tape recorder, with one output connected to the solenoid via a relay amplifier and one to a loudspeaker.

The method of preparing the tapes was as follows. Each instruction used in the experiment was recorded onto a master tape. From this tape a number of copies were made. Half of these (Type A) were recorded with the bottom track left clean and half (Type B) with a continuous square wave of 600 cps on this track. The first instruction was cut out of a Type A copy and the second instruction out of a Type B copy. These were then spliced into clean tape in such a way that the time between the beginning of the first instructions and beginning of the second instructions was 2 sec. This method ensured that the pulse which operated the tachistoscope coincided with the start of the second instruction.

On each trial a neutral instruction was given in addition to the critical instruction. The neutral instruction was either "right" or "left" and told *S* on which side of the answer slip he was to record his response. With two possible critical instructions, two possible neutral instructions, and two possible orders for the critical and neutral instructions, there were eight alternatives under each condition. A separate control tape was prepared for each of these alternatives.

All *S*s were tested individually. Alternatives in which the critical instruction came first and those in which this instruction came second were given alternately. Each *S* received eight practice trials, with each alternative represented once. This was followed in immediate succession by three sets of eight trials, with each alternative occurring once in each set. Subject to these constraints, the alternatives were given in a random order which was different for each *S*.

The 48 volunteer *S*s were graduate and undergraduate students at Birkbeck College and were unaware of the purpose of the experiment. Twelve of these *S*s were tested under each condition. Certain potential *S*s were rejected: *S*s who were unable to carry out the instructions correctly and consistently, and, *S*s who performed without error on the last four practice trials. Presumably all eight stimuli were within the perceptual span of these *S*s. The task was chosen to be not too far beyond the ability of *S*s in order to encourage the use of a central fixation point under all conditions.

Results

Two scores were obtained for each *S*: (1) an *a* score, for the number of individual stimuli correctly reported (in their correct positions) under the four alternatives of each condition where the critical instruction was given first; and (2) a *b* score, for the

TABLE 2

Percentage Accuracy of Report

Instruction	Condition			
	I	II	III	IV
Critical first (*a*)	91	69	80	53
Neutral first (*b*)	75	56	81	58
($a - b$)	16	13	−1	−5

numbers correctly reported under the four alternatives where the critical instruction was given second. The mean *a* scores and *b* scores are shown in Table 2 expressed as percentages. It will be seen that the difference $(a - b)$ was $+16\%$, $+13\%$, -1% and -5% under Cond. I, II, III, and IV, respectively, i.e., the accuracy of report was considerably greater when the critical instruction was given before presentation under the first two conditions but not under the second two conditions.

It was decided to analyze whether the first of these differences was significant separately from the other three, as Cond. I differed from the remaining conditions in providing more than one basis for the selection of the required stimuli. Accordingly, a *t* test was performed to discover whether the mean difference of $+16\%$ differed significantly from zero. The value of *t* was 4.1 ($df = 11$, $P < .01$). An analysis of variance was then performed to test whether the mean values of $(a - b)$ differ significantly from zero for Cond. II, III, and IV, and also whether they differ between conditions. The variability was approximately the same in the three independent groups of 12 values of $(a - b)$. The *F* ratio for Conditions was 9.5 (2 and 33 *df*, $P < .001$). This shows that the effect of the temporal positioning of the critical instruction differed significantly between the conditions. A *t* test was then performed to see whether the mean value of $(a - b)$ for each condition differed significantly from zero, with the *SE* calculated from the within-conditions variance. The values of *t* were 4.4, 0.5 and 1.3 for Cond. II, III, and IV, respectively. The first of these values is significant at beyond the .01 level but the other two are close to the value of unity expected on the null hypothesis. This shows that when the critical instruction was given before presentation, it increased the accuracy of report under Cond. II, as it did under Cond. I, but that no significant effect was observed under Cond. III and IV.

DISCUSSION

The specific object of the experimental design was to eliminate the opportunity for differential effects during retention as a function of whether the critical instruction was given first or second. A potential weakness of the design is that the simultaneous presentation of the critical instruction with the stimulus field might interfere with perception more than the simultaneous presentation of the neutral instruction. If this were so, positive results—in the sense of more accurate report when the critical instruction was given first—would be attributable to the lessened interference with perception as a result of having the neutral instead of the critical instruction simultaneous with perception: they would not show that having the critical instruction before presentation makes possible a selective process during perception. Fortunately this possibility seems unlikely on the basis of the actual results obtained. For if the critical instruction produced more interference with perception than the neutral instruction, positive results should have been obtained in all four experiments unless the ad hoc assumption is made that different critical instructions interfered with perception to differing extents. In fact, however, positive results were obtained only under Cond. I and II and were not obtained under Cond. III and IV. It therefore seems safe to conclude that Cond. I and II demonstrate that a selective process can operate during perception when *S* knows which of two sets of stimuli he is to attempt to report.

The common feature of Cond. I and II was that the critical instruction told

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

EVIDENCE FOR A SELECTIVE PROCESS DURING PERCEPTION OF TACHISTOSCOPICALLY PRESENTED STIMULI[1]

JOHN BROWN

Birkbeck College, University of London

It has usually been found that, when a complex stimulus field is presented tachistoscopically, *S* will report features of the field specified in an instruction more accurately if the instruction is given immediately before rather than immediately after presentation (Vernon, 1952). From Külpe onwards it has usually been concluded that, when the instruction is given before presentation, a selective set is established which leads to more efficient perception. An alternative interpretation would be that the set has influenced not the perceptual process itself but the concurrent storage of the perceived information without which subsequent report could not take place: This will be considered further in the discussion.

For the moment we may confine ourselves to the question whether a selective process can operate *during* perception of a tachistoscopically presented stimulus field irrespective of whether it operates *within* perception. There are two reasons for being dissatisfied with the existing evidence. First, in certain recent experiments no evidence for a selective process during perception was obtained (Augenstine, Blank, Quastler, & Wayner, 1956; Lawrence & Coles, 1954; Lawrence & Laberge, 1956). Secondly, an inadequate experimental design has been employed. It has been the custom to make a simple comparison between the accuracy of report when the instruction immediately precedes presentation and the accuracy when it immediately follows presentation. Greater accuracy under the former condition has been assumed to reveal the operation of a selective process during perception. This is to overlook the possibility that differential effects during retention have been responsible for the increase in accuracy. When the instruction is given before presentation, *S* can make his report immediately and without interruption. When it follows presentation, however, *S* has to attend for it and perceive it before he can record his report. Moreover, any short gap between presentation and the arrival of the instruction will be a gap during which *S* will not know which features of the field to rehearse. Thus higher recall when the instruction comes before presentation might be due to more favorable retention conditions rather than to a selective process during perception, especially as there is evidence that retention in immediate memory is very easily disturbed (Brown, 1958).

In order to avoid differential effects during retention, a more elaborate experimental design is needed. The design adopted in the present experiment involves two instructions to *S* on each trial, the second of which is simultaneous with presentation of stimuli. A "critical" instruction told him what he was to attempt to report and a "neutral" instruction (i.e.

---

[1] The research reported in this paper was supported by a grant from the Central Research Fund, University of London.

neutral with respect to the perceptual process) told him on which side of an answer form he was to record his report. Under one condition the critical instruction was given simultaneously with presentation; under a second condition the order of the instructions was reversed. With this design an instruction is given at the time of presentation under both conditions and there is no interval between presentation and the arrival of the second instruction. Thus the design eliminates the opportunity for differential effects during retention and any difference between the conditions will be attributable to the presence or absence of a selective process during perception. A presupposition of the design is that it takes time to perceive the critical instruction and further time to establish the necessary set for a selective process to operate: otherwise, a selective process would be possible even when the critical instruction was simultaneous with perception.

Method

Two sets of stimuli were presented on each trial. The *S*'s task was to report the set specified in the critical instruction. Under Cond. I this instruction enabled *S* to select the required set on the basis of its class (digits versus consonants), position, and color. Under Cond. II the required set could be selected only by its class, under Cond. III only by its position, and under Cond. IV only by its color.

The eight stimuli for each trial were stenciled on to a white card and were arranged in two rows of four, one immediately below the other. The four outer stimuli, i.e., the stimuli at either end of the two rows, formed one set and the four inner stimuli formed the second set. Thus a fixation point in the center was appropriate whichever set of stimuli *S* was required to report. The stimuli (height 5 mm.) of each set were either all consonants or all digits (selected from the digits 0 to 9) and were either all red or all black. Under Cond. I, the outer set was black consonants and the inner set was red digits. The critical instruction was either "letters" or "digits" and determined for *S* which set he attempted to report. It will be seen that selection of the correct set was possible on the basis of its class, its position (outer or inner), or its color. Under Cond. II all the stimuli were black and two alternative arrangements were used equally often and in random order. In one arrangement the outer set was consonants, and the inner set was digits. In the second arrangement the positions of the consonants and the digits were exchanged. The critical instruction was again either "letters" or "numbers" but here it indicated only the class of the required set, since the position of the consonants and digits was varied and they were all the same color. Under Cond. III the stimuli were the same as for Cond. II. However, the critical instruction was now either "outer" or "inner" so that selection of the required set was possible only on the basis of its position. Under Cond. IV, all the stimuli were consonants and two alternative arrangements were used, again with equal frequency and in random order. In one arrangement, the outer set was red and the inner set was black, while in a second arrangement the positions of the colors were reversed. The critical instruction was either "red" or "black" and selection of the required set was therefore possible only on the basis of its color. Where alternative arrangements were used, as in Cond. II, III, and IV, each arrangement was used as often when the critical instruction preceded as when it followed the neutral instruction (defined below). Cards for 32 trials under each condition were prepared. The stimuli were randomly chosen for each card except that no consonant or digit was permitted to occur more than once. The differences between the conditions are readily seen if Table 1 is examined.

The tachistoscope consisted essentially of a camera shutter and lens which projected a natural-size image of the illuminated card bearing the stimuli onto a ground glass screen. The *S* viewed the screen from the opposite side at whatever he found to be his comfortable reading distance—usually about 40 cm. The screen was masked by a cardboard rectangle except for a 4 × 4-cm. aperture in which the stimuli appeared. Four lines from the corners of this cardboard pointed to the center of the aperture in order to assist *S* in maintaining the correct fixation. The exposure time was .09 sec. (nominal

TABLE 1

Stimulus Materials and Critical Instructions

Cond.	Set	Color	Class	Critical Instructions
I	outer	B (black)	C (consonants)	"letters" or "numbers"
	inner	R (red)	D (digits)	
II	*either*			"letters" or "numbers"
	outer	B	C	
	inner	B	D	
	or			
	outer	B	D	
	inner	B	C	
III	same as Cond. II			"outer" or "inner"
IV	*either*			"red" or "black"
	outer	R	C	
	inner	B	C	
	or			
	outer	B	C	
	inner	R	C	

shutter speed .1 sec.). The room was dimly lit and *S* recorded his report immediately after each presentation on a paper slip with a vertical line down the center, which *E* then removed. The tachistoscope was operated by a solenoid controlled by a pulse recorded on the bottom track of a length of magnetic tape. The top track of this control tape carried the two instructions (see below). The tape was fed through a two-channel tape recorder, with one output connected to the solenoid via a relay amplifier and one to a loudspeaker.

The method of preparing the tapes was as follows. Each instruction used in the experiment was recorded onto a master tape. From this tape a number of copies were made. Half of these (Type A) were recorded with the bottom track left clean and half (Type B) with a continuous square wave of 600 cps on this track. The first instruction was cut out of a Type A copy and the second instruction out of a Type B copy. These were then spliced into clean tape in such a way that the time between the beginning of the first instructions and beginning of the second instructions was 2 sec. This method ensured that the pulse which operated the tachistoscope coincided with the start of the second instruction.

On each trial a neutral instruction was given in addition to the critical instruction. The neutral instruction was either "right" or "left" and told *S* on which side of the answer slip he was to record his response. With two possible critical instructions, two possible neutral instructions, and two possible orders for the critical and neutral instructions, there were eight alternatives under each condition. A separate control tape was prepared for each of these alternatives.

All *S*s were tested individually. Alternatives in which the critical instruction came first and those in which this instruction came second were given alternately. Each *S* received eight practice trials, with each alternative represented once. This was followed in immediate succession by three sets of eight trials, with each alternative occurring once in each set. Subject to these constraints, the alternatives were given in a random order which was different for each *S*.

The 48 volunteer *S*s were graduate and undergraduate students at Birkbeck College and were unaware of the purpose of the experiment. Twelve of these *S*s were tested under each condition. Certain potential *S*s were rejected: *S*s who were unable to carry out the instructions correctly and consistently, and, *S*s who performed without error on the last four practice trials. Presumably all eight stimuli were within the perceptual span of these *S*s. The task was chosen to be not too far beyond the ability of *S*s in order to encourage the use of a central fixation point under all conditions.

Results

Two scores were obtained for each *S*: (1) an *a* score, for the number of individual stimuli correctly reported (in their correct positions) under the four alternatives of each condition where the critical instruction was given first; and (2) a *b* score, for the

TABLE 2

Percentage Accuracy of Report

Instruction	Condition			
	I	II	III	IV
Critical first (*a*)	91	69	80	53
Neutral first (*b*)	75	56	81	58
(*a* − *b*)	16	13	−1	−5

numbers correctly reported under the four alternatives where the critical instruction was given second. The mean *a* scores and *b* scores are shown in Table 2 expressed as percentages. It will be seen that the difference $(a - b)$ was $+16\%$, $+13\%$, -1% and -5% under Cond. I, II, III, and IV, respectively, i.e., the accuracy of report was considerably greater when the critical instruction was given before presentation under the first two conditions but not under the second two conditions.

It was decided to analyze whether the first of these differences was significant separately from the other three, as Cond. I differed from the remaining conditions in providing more than one basis for the selection of the required stimuli. Accordingly, a *t* test was performed to discover whether the mean difference of $+16\%$ differed significantly from zero. The value of *t* was 4.1 ($df = 11$, $P < .01$). An analysis of variance was then performed to test whether the mean values of $(a - b)$ differ significantly from zero for Cond. II, III, and IV, and also whether they differ between conditions. The variability was approximately the same in the three independent groups of 12 values of $(a - b)$. The *F* ratio for Conditions was 9.5 (2 and 33 *df*, $P < .001$). This shows that the effect of the temporal positioning of the critical instruction differed significantly between the conditions. A *t* test was then performed to see whether the mean value of $(a - b)$ for each condition differed significantly from zero, with the *SE* calculated from the within-conditions variance. The values of *t* were 4.4, 0.5 and 1.3 for Cond. II, III, and IV, respectively. The first of these values is significant at beyond the .01 level but the other two are close to the value of unity expected on the null hypothesis. This shows that when the critical instruction was given before presentation, it increased the accuracy of report under Cond. II, as it did under Cond. I, but that no significant effect was observed under Cond. III and IV.

DISCUSSION

The specific object of the experimental design was to eliminate the opportunity for differential effects during retention as a function of whether the critical instruction was given first or second. A potential weakness of the design is that the simultaneous presentation of the critical instruction with the stimulus field might interfere with perception more than the simultaneous presentation of the neutral instruction. If this were so, positive results—in the sense of more accurate report when the critical instruction was given first—would be attributable to the lessened interference with perception as a result of having the neutral instead of the critical instruction simultaneous with perception: they would not show that having the critical instruction before presentation makes possible a selective process during perception. Fortunately this possibility seems unlikely on the basis of the actual results obtained. For if the critical instruction produced more interference with perception than the neutral instruction, positive results should have been obtained in all four experiments unless the ad hoc assumption is made that different critical instructions interfered with perception to differing extents. In fact, however, positive results were obtained only under Cond. I and II and were not obtained under Cond. III and IV. It therefore seems safe to conclude that Cond. I and II demonstrate that a selective process can operate during perception when *S* knows which of two sets of stimuli he is to attempt to report.

The common feature of Cond. I and II was that the critical instruction told

S the class of the required stimuli. Under Cond. III it told *S* their position only, and under Cond. IV, their color only. The results thus provide evidence that a selective process can operate during perception when the class of the required stimuli is known but they fail to provide any evidence that such a selective process is possible when only the position or only the color is known. The results do however contain one rather puzzling feature. Although the same stimulus material was used under Cond. II and III, Table 2 shows that the *b* score (the score when the critical instruction was second) under Cond. II was considerably below the *b* score under Cond. III and a *t* test showed that the difference is significant at the .02 level. This might be taken to show that selection of the correct set of stimuli is inherently easier on the basis of position so that, when the critical instruction specifies the position, it is unnecessary for it to precede presentation. But if this interpretation is accepted it becomes difficult to explain why it is an advantage to have the critical instruction before presentation under Cond. I, since under this condition the critical instruction informed *S* not only of the class but also of the position of the required stimuli. An alternative interpretation of the difference between *b* scores under Cond. II and III is that they are due to a chance difference between the Cond. II and the Cond. III *S*s. Such a difference might affect the validity of comparing $(a - b)$ scores, specifically if the difference score $(a - b)$ of each *S* were correlated with his *b* score. However, examination of the relation between $(a - b)$ scores and *b* scores within each condition failed to reveal any consistent relationship between the two.

Conditions I and II demonstrate a selective process operating during perception but this process is not necessarily within perception itself. When a complex stimulus field is presented tachistoscopically, *S*'s report consists of several responses made one after another and only the first of these, if any, can be regarded as immediate in a literal sense. The efficiency of report must therefore depend partly on the efficiency with which the stimulus information is stored as it is perceived. It is possible, therefore, that the selective process concerns storage, i.e., learning, rather than perception. A previous experiment seemed to show that a selective effect on storage is indeed possible (Brown, 1954). Unless or until contrary evidence is forthcoming, it would therefore be more economical to suppose that the present results are also due to a selective effect on learning and not to an effect on perception.

SUMMARY

The experiment investigates whether and on what bases a selective process can operate during the perception of a briefly exposed stimulus field. The field, which contained two sets of stimuli, was presented tachistoscopically for .09 sec. The *S*'s task was to report the set of stimuli specified in a "critical" instruction in a place specified in a "neutral" instruction. Under one alternative, the "critical" instruction was given 2 sec. before the stimulus field and the "neutral" instruction simultaneously with the field. Under a second alternative, the order of the two instructions was reversed. Forty-eight *S*s were tested, 12 under each of four conditions.

When the critical instruction was given in advance, *S* knew the following properties of the required set of stimuli at the time of presentation: under Cond. I, their position, color, and class; under Cond. II, their class; under Cond. III, their position; under Cond. IV, their color. Accuracy of report was significantly higher with this alternative under Cond. I and II but not under Cond. III and IV. It is concluded that a selective process which increased the accuracy of report was able to operate during perception on the basis of class but was not able to operate on the basis of position or color. Further work is needed to show whether this selective process acted in perception itself or in the concurrent learning process.

REFERENCES

AUGENSTINE, L., BLANK, A. A., QUASTLER, H., & WAYNER, M. Human performance in information transmission: IV. Flash recognition of familiar displays. Control systems lab., Univer. of Illinois, Contract DA-36-039-SC-56695, Rep. R69, 1956.

Brown, J. The nature of set-to-learn and of intramaterial interference in immediate memory. *Quart. J. exp. Psychol.*, 1954, **6,** 141–148.

Brown, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10,** 12–21.

Lawrence, D. H., & Coles, G. R. Accuracy of recognition with alternatives before and after the stimulus. *J. exp. Psychol.*, 1954, **47,** 208–214.

Lawrence, D. H., & Laberge, D. L. Relationship between recognition accuracy and order of reporting simulus dimensions. *J. exp. Psychol.*, 1956, **51,** 12–18.

Vernon, M. D. *A further study of visual perception.* Cambridge, Eng.: Cambridge Univer. Press, 1952.

(Received April 4, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

EFFECT OF NUMBER OF DOUBLETS UPON VERBAL MAZE LEARNING

JAMES F. VOSS[1] AND J. A. ZIEGLER

Wisconsin State College at Eau Claire

Results of a previous experiment (Thompson, Voss, & Brogden, 1957) on maze patterning indicated that in a four-alternative, 16-unit verbal maze, the presentation of a doublet (two successive like responses) yielded a significant increase in errors at the second position of the doublet, regardless of the doublet locus. Since a first trial error analysis revealed that *S*s tended not to repeat the previously correct response, the findings were considered in terms of the guessing tendencies of *S*s and the maze pattern. Moreover, since guessing tendencies are likely maximal early in acquisition and maze pattern effects likely increase as the initial guessing habits diminish, the results suggest that the doublet effect is a differential function of two stages of acquisition, previously designated discovery and fixation (Melton, 1950).

The doublet effect is probably limited to a specific number of doublets, since use of a complete list of doublets would likely facilitate discovery of the pattern and eliminate the doublet effect. Moreover, although acquisition was not significantly influenced by the presence of a doublet (Thompson, Voss, & Brogden, 1957), facilitation would be expected with an entire list of doublets because of the probable rapid pattern discovery which would effectively shorten the sequence by one-half.

The purpose of the present study was to investigate the doublet effect and acquisition as a function of number of doublets, and to determine the relative role of discovery and fixation in these relations. Discovery was defined as the number of trials to the first correct response at a given serial position and fixation was defined as the number of trials from the first correct response to the correct response after which no further errors occurred at the particular serial position.

METHOD

One of four numbers, 10, 20, 30, and 40, was correct at each position in a 16-unit verbal maze. Each number occurred four times in one maze pattern. The control group pattern contained no doublets or split-doublets (e.g., 20–10–20) and was identical to the control group pattern employed in the study of Thompson, Voss, and Brogden (1957). For the five experimental groups, the control maze pattern was changed to provide a doublet (10–10, 20–20, 30–30, or 40–40) for Maze Units 6 and 7 in Group E_1; Units 5 and 6, 9 and 10 in Group E_2; Units 3 and 4, 7 and 8, 11 and 12, 13 and 14 in Group E_3; Units 3 and 4, 5 and 6, 7 and 8, 9 and 10, 11 and 12, 13 and 14 in Group E_4; and all pairs of maze units in Group E_5. The selection of doublet positions for Groups E_1, E_2, and E_3 was random with the limitation that no doublets occur at Maze Units 1 and 2 or 15 and 16. The pattern for each one of the experimental groups and that of the control group was permuted as in the Thompson, Voss, and Brogden (1957) study in order to provide for the representation of each number once in each of the 16 positions.

The control group and each experimental group contained 24 *S*s. Six *S*s were assigned to one of the four subpattern permutations of each group. Each of two *E*s handled half of the *S*s and *E* was orthogonal to the other experimental conditions. The criterion of learning was one errorless trial. The *S*s were volunteers from Wisconsin State College at Eau Claire and were assigned randomly to the various conditions. The instructions and other experimental conditions were analogous to those of Thompson, Voss, and Brogden (1957).

---

[1] Now at the College of Wooster.

RESULTS

The first trial data for all groups were analyzed to determine whether first trial errors increase as a function of number of doublets (Thompson, Voss, & Brogden, 1957). The Trial 1 mean errors for the Control Group was 20.0 and for Groups E_1 to E_5 were: 21.0, 21.9, 24.0, 23.6, and 25.0. An analysis of variance on the four permutations revealed no significant differences. The permutation data were pooled and a subsequent analysis of variance revealed that the number of doublets source of variation was significant (.01 level). Duncan range tests (Duncan, 1951) revealed that each group differed significantly from each other group with the exception of Groups E_3 and E_4.

Figure 1 presents the number of trials, total errors, and total time to criterion as a function of number of doublets. An analysis of variance on the four permutations for the five experimental groups and the control group revealed no significant permutation effect for any of the three measures. The permutation data were therefore pooled and subsequent analyses of variance revealed the number of doublets source of variation was significant (.01 level) for all three measures of acquisition. The E and E × Number of doublets sources of variation were not significant. Results of Duncan range tests performed on the number of doublet means are as follows:[2] trials: $E_2CE_1E_3E_4E_5$; time: $E_2E_1E_3CE_4E_5$; errors: $E_2E_3E_1CE_4E_5$.

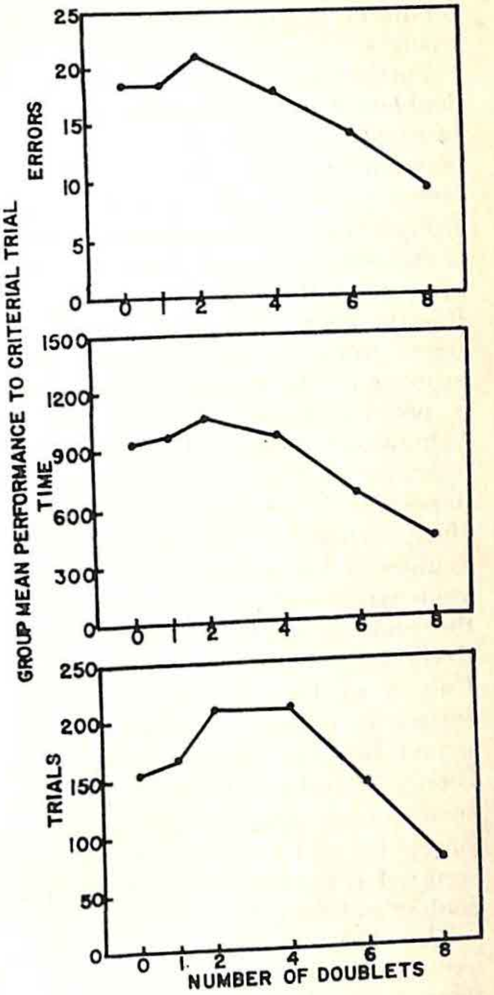


FIG. 1. Group mean performance to criterial trial as a function of number of doublets.

Analysis of variance of serial position data of the experimental groups, corrected for serial position per se by subtraction of the control group serial position means, revealed significant Serial position × Number of doublets and Serial position × Number of doublets × Permutations sources of variation. The former result indicates that serial position differences are a function of the doublet condition or concomitant sequential changes.

---

[2] Groups underlined by a continuous line do not differ significantly from each other; groups not included in one underline differ significantly from those underlined.

Further analysis of the effect of the doublets was performed by applying two-tailed *t* tests of the difference between means (difference scores between experimental and control groups) of the first and second member of the doublet over all of the doublets in each of the experimental groups. Results show that the increase in errors from the first to the second member of the doublet in Group E_1 is not significant. Previous results (Thompson, Voss, & Brogden, 1957) indicated that serial position was not important in obtaining the doublet effect. However, no condition had a doublet at Units 6 and 7 in the previous study and the control group of the previous study also had a relatively large increase in errors from Unit 6 to Unit 7. Significant increases in errors are found for the second member of the doublets in Groups E_2 and E_3 (.05 level), whereas no significant differences occurred for Group E_4, and a significant decrease occurred at the second member of the doublet in Group E_5 (.01 level).

The discovery and fixation data were analyzed over all doublet positions for each group. The *t* tests indicated a significant increase in the number of discovery trials between the first and second members of the doublet for Groups E_2 and E_3 and a significant decrease in fixation trials in Groups E_1, E_3, and E_5.

DISCUSSION

The first trial error analysis corroborates the previous findings (Thompson, Voss, & Brogden, 1957) that indicated the doublet effect is a function of guessing tendencies.

The serial position and doublet analyses indicate that the doublet effect occurs to four doublets, with a transition phase at six doublets, and a reversal of the doublet effect at eight doublets. Moreover, the discovery and fixation results, when considered with the over-all doublet analysis further supports the findings (Ernst, Hoffeld, Seidenstein, & Brogden, 1960) which attribute the doublet effect to the discovery phase of learning.

The acquisition analyses reveal that an increase from one to two or four doublets inhibited acquisition (nonsignificant), whereas eight doublets facilitated acquisition. Comparison of the acquisition results with the other findings reveals that the inhibitory factors in acquisition are associated with a significant doublet effect, primarily attributable to discovery. However, the facilitation of acquisition in Group E_5 is associated with reversal of the doublet effect and more rapid fixation.

SUMMARY

Acquisition of a four-alternative, 16-choice-point verbal maze was studied as a function of 0, 1, 2, 4, 6, or 8 doublets. The results indicated that first trial errors increase as a function of number of doublets, that the previously found doublet effect is a function of the discovery stage of learning, and that acquisition varied as a function of number of doublets.

REFERENCES

DUNCAN, D. B. A significant test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2**, 171–189.

ERNST, R. L., HOFFELD, D. R., SEIDENSTEIN, S., & BROGDEN, W. J. Relation of serial position errors to doublet and split-doublet location in verbal maze pattern. *J. exp. Psychol.*, 1960, **59**, 94–103.

MELTON, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* New York: Macmillan, 1950. Pp. 668–690.

THOMPSON, R. F., VOSS, J. F., & BROGDEN, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, 1957, **54**, 253–258.

(Received April 8, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

THE RELATIONSHIP OF TEST AND GENERAL ANXIETY, DIFFICULTY OF TASK, AND EXPERIMENTAL INSTRUCTIONS TO PERFORMANCE

IRWIN G. SARASON AND ERNEST G. PALOLA[1]

University of Washington

Following upon the development of questionnaire techniques for assessing anxiety in *S*s, at least two lines of investigation have become clearly discernible. One line has studied the reactions of *S*s differing in anxiety scores to situations which may be considered as threatening or stressful (Child, 1954; Sarason, 1956, 1957a, 1957b). The other line has studied the relationship of anxiety scores to task characteristics, e.g., the association value and intralist similarity of material to be learned (Farber, 1955; Montague, 1953; Spence, 1958; Taylor, 1956). Although each line of investigation has produced some consistent and interesting findings, little work has been done in the direction of interrelating them. The three experiments reported in the present paper represent attempts in this direction.

Studies on the relationship between anxiety and stress have typically assumed that the performance of high-anxious *S*s would be more detrimentally affected by experimentally induced personal threat than would be true for low-anxious *S*s. Under neutral conditions no differences due to anxiety would be expected. Several experiments have provided support for these assumptions and have led to the suggestion that anxiety be conceived as a class of interfering responses which can be evoked by manipulating one set of stimuli presented to *S* in an experiment, the instructions *E* communicates to *S* (Child, 1954; Sarason 1958b). It has also been suggested that the more directly related the anxiety measure is to the experimental instructions, the more meaningful will be the relationships obtained. There is some evidence in support of this expectation (Sarason, 1958b). However, our present knowledge of the relative effects of test and general anxiety is much too limited.

Studies on the relationship between anxiety and task characteristics have typically involved the manipulation of variables which might affect *S*'s response hierarchy. The results of several experiments have supported the hypothesis that the greater the response competition in the hierarchy, the greater the advantage in performance expected of low-anxious as compared with high-anxious *S*s (Farber, 1955; Spence, Farber, & McFann, 1956; Taylor 1956).

It thus would appear that there is evidence supporting a relationship between (*a*) anxiety and stress and (*b*) anxiety and task characteristics. Despite these two sets of interesting findings a number of unresolved issues becloud the picture.

In experiments on the interaction of anxiety and experimentally induced threat there is the problem of the effect on threat arousal of the face validity for *S* of the task employed. On the other hand, studies of anxiety–

---

[1] This research was supported in part by a grant (M-2397) from the National Institute of Mental Health.

task relationships have not systematically investigated the interaction of these variables with the motivational properties of the experimental instructions.

In each of the three experiments reported here three variables were studied: (*a*) anxiety, (*b*) task difficulty, and (*c*) differential motivating instructions. Two measures of anxiety, one general and one specific to *S*'s reactions in testing situations, were employed. Because of the use of an evaluational threat it was predicted that the latter measure would relate more systematically to performance in these experiments than would the former measure.

What will, perhaps, be of most interest in the present experiments is the possibility of the emergence of triple interactions between anxiety, instructions, and difficulty. Should triple interactions appear, the necessity of reformulating and tying more closely together the various interpretations of anxiety would seem indicated.

EXPERIMENT I

Method

Subjects. The *S*s were 121 students in introductory psychology classes at the University of Washington. Prior to and independent of the experimental situation, *S*s had been administered the Test Anxiety Scale (TAS) (Sarason, 1958a), the Bendig short form Manifest Anxiety Scale (MAS) (Bendig, 1956; Taylor, 1951), and the Social Desirability Scale (SDS) (Edwards, 1957).

Procedure. Each *S* performed on either an easy or difficult task. The words easy and difficult do not imply any absolute denotation but rather refer to the relative difficulty of the two tasks. The easy task was an expanded version of the digit–symbol subtest of the Wechsler-Bellevue Intelligence Scale for Adolescents and Adults, Form I. The hard task differed from this subtest only with respect to the symbols which *S* assigned to the code numbers. The nine symbols for the difficult task were variations of the letter L. Each symbol had attached to this L a different flag or hook. Thus all nine symbols differed from each other but these differences were slight and consequently increased the difficulty of making discriminations between symbols. These two forms of the digit–symbols test were administered to groups with a time limit of 3 min. After performing on the experimental tasks *S*s were asked to rate the tasks with respect to their interest value.

Half of the *S*s performed under "IQ" instructions, half under neutral instructions. The IQ instructions were as follows:

> The test you are about to take is part of the Wechsler-Bellevue Intelligence Scale for Adults. This test has been found to predict such things as course grades, success in later life, and to some extent the kind of personality you possess. Of course, your own intelligence will primarily determine whether you do well or poorly on this test. At a later date there will be an opportunity for each of you to compare your IQ with those of the other people in this group. You will then be able to determine how your abilities and capacities compare with other people like you.

The neutral instructions were:

> We are conducting a research project on how interesting different kinds of tests and tasks are. After completing the test I am about to give you, I'm going to ask you to rate the test in terms of how you enjoyed it. I want to make it clear that this is an experimental test and has nothing to do with any course offered at the University of Washington. We appreciate this contribution which you are making to a better understanding of the psychology of individual preferences.

Since *S*s were not preselected in terms of TAS, MAS, and SDS, it was necessary for statistical purposes to randomly exclude certain *S*s from the analyses of variance involving these three variables. The *N* for the TAS analysis was 104, for the MAS analysis, 108, and for the SDS analysis, 104.[2] The *S*s were divided on these variables at the median. High-TAS *S*s had scores of 6 and above; low-TAS *S*s, scores of 5 and below;

---

[2] In each of the three experiments reported here, the various groups studied included approximately 50% male *S*s and 50% female *S*s.

TABLE 1

MEANS AND *SD*s OF NUMBER OF CORRECT RESPONSES FOR TEST ANXIETY GROUPS ON DIGIT–SYMBOL TEST (EXP. I)
(N = 15 for IQ Groups, N = 11 for Neutral Groups)

Test Anxiety	Difficult Form				Easy Form			
	IQ		Neutral		IQ		Neutral	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
High—	65.60	10.61	70.64	13.88	123.20	13.47	109.45	12.09
Low—	81.27	12.86	62.18	17.01	110.73	18.10	101.00	22.83

high-MAS *S*s had scores of 4 and above; low-MAS *S*s had scores of 3 and below. High-SDS *S*s had scores of 33 and above; low-SDS *S*s had scores of 32 and below. It is clear that the designations of high and low denote relative and not absolute standing on these personality dimensions.

Results

The data were analyzed by means of three 2 × 2 × 2 analyses of variances. The variables were: (*a*) high and low scores on each of the three scales used, (*b*) difficult and easy digit–symbol tasks, and (*c*) IQ and neutral instructions.

TAS. Table 1 presents the means and *SD*s for the eight experimental groups. Table 2 presents a summary of the analysis of variance. As expected, the F for difficulty of task was highly significant ($P < .001$). The groups performing under the IQ instructions performed at a significantly higher level than did the neutral *S*s ($P < .005$). The TAS × Difficulty interaction ($P < .025$) reflects a superiority of low-TAS *S*s to high-TAS *S*s on the difficult digit–symbol test and a superiority of high- to low-TAS *S*s on the easy task. Although an acceptable level of significance was not obtained for TAS × Instructions, there was a suggestion of a superiority in performance of low- to high-TAS *S*s under the IQ condition, and the reverse was the case under the neutral condition ($.10 < P < .20$). A significant F for the triple interaction, TAS × Instructions × Difficulty ($.025 < P < .05$) was obtained. Further study of the means in Table 1 revealed that this triple interaction reflected an interaction between TAS and Instructions for the difficult digit–symbol task but not for the easy task. For the easy task, high-TAS *S*s performed at higher levels than did the low-TAS *S*s for both the IQ and neutral conditions. On the difficult task this superiority of high- to low-TAS *S*s held only for the neutral condition, with the low-TAS *S*s performing at a higher level than the high-TAS *S*s

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT RESPONSES FOR EIGHT TEST ANXIETY GROUPS ON DIGIT–SYMBOL TEST (EXP. I)

Source	*df*	*F*	*P*
Test anxiety (TAS)	1	.7193	
Difficulty (D)	1	176.2386	.001
Instructions (I)	1	8.7775	.005
TAS × D	1	6.7270	.025
TAS × I	1	2.5204	
D × I	1	.5544	
TAS × I × D	1	4.9330	.05
Within groups error (MS)	96	(254.55)	
Total	103		

under the IQ condition. These results, no doubt, suggest in part the basis for failure of the TAS × Instructions effect to attain significance.

MAS. A 2 × 2 × 2 analysis of variance of the MAS data revealed only two significant effects. In addition to a significant F (189.53) for difficulty of task ($P < .001$), there was a significant F (13.13) for Instructions ($P < .001$). As was true in the TAS analysis, the groups receiving the IQ instructions performed at higher levels than the groups receiving the neutral instructions. None of the interactions obtained was significant, although there was a suggestion ($.10 < P < .20$) of an interaction between MAS scores and task difficulty, with the direction of the means similar to those obtained in the earlier TAS × Difficulty interaction.

SDS. An analysis of variance of the SDS data revealed significant Fs for Difficulty ($P < .001$) and Instructions ($P < .005$). The direction of the differences among the means was the same as in the TAS and MAS analyses. The only interaction F which approached significance was that for SDS × Difficulty × Instructions ($P < .10$).

Discussion

As predicted, TAS was more significantly related to performance than was MAS. The possibility of an interaction between the anxiety and difficulty variables was confirmed only for the TAS measure.

The results for TAS × Instructions and the TAS × Difficulty × Instructions suggest that the interaction between test anxiety and instructions held only for those *S*s who performed on the difficult task. The significant triple interaction provides a good basis for the contention that relating anxiety to either difficulty of task or experimental instructions will lead to the discovery of relationships not nearly as fruitful or meaningful as relating anxiety to both variables. On the basis of this study one might infer that both a difficult task and highly motivating conditions increase the emission of interfering anxiety responses for highly anxious *S*s as defined by TAS.

EXPERIMENT II

The main outlines of this experiment followed closely those of Exp. I with the study of the same three variables: anxiety (high and low), difficulty of task (easy and hard), and experimental instructions (IQ and neutral). However, there were task and procedural differences.

Method

Subjects. The *S*s were students in introductory psychology and introductory sociology at the University of Washington. Prior to participation in the experiment proper a large number of students had been administered the TAS and MAS. From these students those in approximately the upper and lower 20% of the TAS and MAS distributions were selected as *S*s. There were 80 *S*s (10 per experimental group) in the TAS and MAS analyses of variance. High-TAS *S*s had scores of 9 or higher; low-TAS *S*s had scores of 2 or below. High-MAS *S*s had scores of 10 or higher; low-MAS *S*s had scores of 2 or below.

Procedure. The task was the digit–symbol test administered in groups. The difficult form was identical to the one described in Exp. I. The easy form used letters of the alphabet as symbols rather than the symbols used in the comparable form of Exp. I. The *S*s performed for 5 min. on the digit–symbol test.

The IQ instructions were:

> The test you are about to participate in is part of the Wechsler Intelligence Scale for Adults. Your task is to perform as well as you possibly can. After the results have been compiled you will be given the opportunity to compare your performance on this abbreviated intelligence test to the performance of others in this group.

The neutral instructions were:

> The task you are about to perform is related to a research project conducted on individual differences here at the University of Washington. We are simply interested in the value of the particular tests you will take for further research. We wish to thank you for your cooperation in this research.

Results

Only two effects in the 2 × 2 × 2 analyses of variance involving the TAS were statistically significant. The F for Difficulty was 168.83 ($P < .001$), and the triple interaction F, TAS × Difficulty × Instruction, was 8.52 ($P < .01$). None of the other effects approached significance.

None of the Fs in the 2 × 2 × 2 analysis of variance involving the MAS approached significance.

Discussion

The significant TAS effects for Difficulty and TAS × Difficulty × Instructions are comparable to the results in Exp. I. However, there were many more significant results in Exp. I than in Exp. II. Perhaps the most important differences between the two experiments, which might explain these different results, are the instructions used. The IQ instructions of Exp. I appear to be more challenging and threatening than do the Exp. II IQ instructions. In the latter set of instructions no statement was made concerning a direct relationship between intelligence and achievement.

Although this instructional difference appears to be most striking in accounting for certain of the discrepant results of Exp. I and Exp. II there are other obvious possibilities, such as experimenter and sampling differences and differences in the easy tasks used. However, despite these factors, the fact of the consistency in the two studies with respect to the Anxiety × Difficulty × Instructions interaction has important implications. As was pointed out earlier, relating anxiety either to the instructional or to the difficulty variable would not appear to provide a sufficiently comprehensive basis for predicting results such as the triple interaction reported here.

EXPERIMENT III

The experimental design for this study was similar to those employed for Exp. I and II. Anxiety defined by TAS and MAS was related to task difficulty (easy and hard tasks) and to experimental instructions (IQ and neutral). The important difference between this experiment and the previous ones was in the nature of the task used. In the present experiment *S*s were called upon to respond to a long list of numbers by indicating to which of several rules the particular numbers adhered. A more detailed description of this task has been presented elsewhere (Palola, 1958). It was hoped that by using a different task in the present experiment it would be possible to gain information about the generality of the previously presented findings.

Method

Subjects. The same procedure for constituting high- and low-TAS and MAS groups was used in Exp. III as was used in Exp. II. The *S*s used in this experiment fell within the upper and lower 20% of the TAS and MAS distribution for a sample of 350 introductory psychology and introductory sociology students. The high and low cutoff points for the TAS and MAS distributions were the same as those used in Exp. II. The TAS and MAS analyses each had an $N = 80$ with 10 *S*s in each of the 8 experimental groups.

Procedure. The task used required *S* to indicate to which of several rules a series of numbers adhered. For the easy task, *S* was instructed to circle all numbers which were divisible by 2 and whose resulting quotient was an even number. For the difficult task seven possible rules had to be considered for each number. These rules were: (1) number

TABLE 3

MEANS AND *SD*s OF NUMBER OF CORRECT RESPONSES FOR TEST ANXIETY GROUPS ON ARITHMETIC TEST (EXP. III)

(N = 10 per group)

Test Anxiety	Difficult Form				Easy Form			
	IQ		Neutral		IQ		Neutral	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
High—	13.0	26.3	12.1	9.7	28.7	28.8	34.2	38.3
Low—	21.1	13.8	19.0	12.2	29.4	39.3	32.2	39.9

is divisible by 2 only; (2) number is divisible by 2 and 5; (3) number is divisible by 3 and 7; (4) number is divisible by 2, 3, 4; (5) number is divisible by 2 and 7; (6) more than one rule applies; (7) none of the rules apply.

The time limit for this task was 5 min. The IQ and neutral instructions were identical to those employed in Exp. II.

Results

TAS. Tables 3 and 4 present respectively the means and *SD*s and a summary of the analysis of variance of the TAS data. Table 4 reveals three significant results. In addition to the expected significant F for Difficulty, the Fs for TAS × Difficulty and TAS × Difficulty × Instructions were found to be significant ($P < .05$ and $P < .01$, respectively). Inspection of Table 3 will reveal that the TAS × Difficulty interaction was determined almost entirely by a superiority of low- to high-TAS *S*s on the difficult task. On the easy task the performances of high- and low-TAS groups were virtually identical.

TABLE 4

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT RESPONSES FOR EIGHT TEST ANXIETY GROUPS ON ARITHMETIC TEST (EXP. III)

Source	*df*	F	P
Test anxiety (TAS)	1	3.7930	
Difficulty (D)	1	71.0690	.001
Instructions (I)	1	.7292	
TAS × D	1	5.3694	.05
TAS × I	1	.3088	
D × I	1	2.5808	
TAS × I × D	1	12.1893	.001
Within groups error (*MS*)	72	(61.85)	
Total	79		

The TAS × Difficulty × Instructions was largely attributable to interaction effects under the difficult condition. Under this condition low-TAS *S*s were superior to high-TAS *S*s. Under the easy condition there was a tendency for *S*s under the neutral condition to perform at a higher level than under the IQ condition.

MAS. Except for the Difficulty variable none of the Fs in the $2 \times 2 \times 2$ MAS analysis of variance was significant.

DISCUSSION

Significant Fs involving anxiety were obtained in each of the three TAS analyses of variance. None of the Fs involving anxiety was significant in the MAS analyses. Since these three experiments involved the manipulation of an evaluational stress condition, the results seem to support the hypothesis that the more closely the anxiety scale specifically taps anxieties aroused in the experimental situation the greater the likelihood of that scale relating to behavior in the specific situation studied.

In our view the most challenging

findings in the present series of experiments were the triple interactions involving test anxiety, task difficulty, and experimental instructions. It seems likely that any comprehensive interpretation of anxiety will have to incorporate and interrelate these variables. The writers would suggest that, at least, the following factors will have to be considered in formulating theoretical and empirical directions in the study of anxiety: (*a*) the instrument used to assess anxiety and the scores obtained by *S*s; (*b*) the nature of the task on which *S* performs (in the present experiments, the nature of the obtained triple interaction was quite different for the arithmetic task than for the digit–symbol task); (*c*) the difficulty level of the task on which *S* performs.

SUMMARY

Three experiments were carried out involving (*a*) anxiety (test anxiety and general anxiety), (*b*) motivational and neutral instructions, and (*c*) difficulty of task. The tasks used were a digit–symbol test and an arithmetic test.

Analyzing the data by means of analyses of variance, significant Test Anxiety × Instructions × Difficulty interactions emerged in each of the three experiments. Test anxiety was related to *S*'s performance more frequently than was general anxiety (as measured by Bendig's short form of Taylor's MAS), and in general it appeared that high difficulty of task and highly motivating instructions combined to affect detrimentally the performance of high-anxious *S*s. It was concluded that attempts at relating anxiety to either the instructional or difficulty variables alone would not be as fruitful as simultaneous analyses of the three variables.

REFERENCES

BENDIG, A. W. The development of a short form of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1956, **20**, 384.

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–171.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.

PALOLA, E. G. Inter-relationships among anxiety, motivation instructions and difficulty of task. Unpublished master's thesis, Univer. of Washington, 1958.

SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.

SARASON, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 166–171. (a)

SARASON, I. G. The effect of anxiety and two kinds of failure on serial learning. *J. Pers.*, 1957, **25**, 383–392. (b)

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344. (a)

SARASON, I. G. Effects on verbal learning anxiety, reassurance, and meaningfulness of material. *J. exp. Psychol.*, 1958, **56**, 472–477. (b)

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

(Received April 9, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

VISUAL PATTERN IDENTIFICATION AS A FUNCTION OF FILL AND DISTORTION [1], [2]

W. A. HILLIX

University of Missouri

The present experiment is concerned with the question, "What characteristics of a stimulus make it difficult for a human observer to identify?" There are two ways of approaching the problem. The first is to study the "absolute" characteristics of stimuli, using a paradigm in which *S* must identify the correct stimulus among others of the same general kind. In this paradigm, the stimulus to be identified is exactly the same as the one studied by *S*. Fitts, Weinstein, Rappaport, Anderson, and Leonard (1956), following Attneave (1954; 1955), have studied within this paradigm. The stimuli used were randomly constructed on various probability models. French (1953; 1954a) has also done similar work, specifically on the relationship between complexity and identifiability. Within this paradigm, the standard stimulus to be identified may either be present or absent; in the latter case, memory may be more or less important depending upon the time intervals involved and the number of stimuli studied before testing. One would expect the simplest relations to obtain where memory is minimally involved.

A slightly different situation is that in which the stimulus which should be chosen as correct differs somewhat from the standard stimulus which is supposed to be matched. One would expect this to be the more common perceptual case, since objects are seldom seen in precisely the same orientation twice, lighting varies, and so on. The paradigm in which the correct stimulus is exactly like the standard is a special case of the more general one, where the value of similarity is maximal.

Attneave (1954) gives some interesting suggestions for quantifying stimulus characteristics. Fitts et al. (1956) address themselves more specifically to the problem of "visual noise"; as more noise is added, the stimulus becomes less similar to its original form. These authors suggest that random rearrangement of the cells within a matrix is one quantifiable method of manipulating similarity. This is the method used in the present study.

French (1954b) previously worked with a problem abstractly quite similar to the present one; he varied the number of dots in a standard pattern and the number of dots added randomly to a version to be recognized. Number of dots in the original pattern and number of "noise" dots added were varied. He found that percentage of error fell as number of dots in the original pattern increased from two to nine and increased as number of noise dots increased from one to eight.

The present experiment used a different range of number of elements in the original pattern, from 10 to 50. The variations were produced by randomly moving different numbers of the original elements, from 10% to 40% of the original number of elements. The *S*'s task was to select that alternative which was most like the standard stimulus from among three stimuli, one of which had been produced by some rearrangement of the standard.

---

[1] This research was supported in part by the United States Air Force under Contract AF 41(657)-11, monitored by the Operator Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

[2] Thanks are due to Robert G. Kinkade for collecting the data of this experiment.

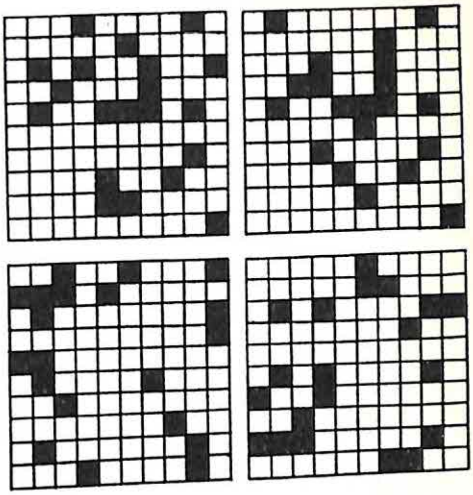


FIG. 1. Sample 20 F, 20 D problem; standard at upper left, correct alternative at upper right.

METHOD

Construction of stimuli. The stimuli were constructed in a square matrix with 10 rows and 10 columns. This gave a matrix of 100 squares, any one of which could be filled or left empty. Stimuli were constructed with five degrees of fill: 10, 20, 30, 40, and 50%. In order to do so, the rows and columns were each numbered from zero through nine so that the square to fill could be determined randomly by selecting a pair of numbers from a table of random numbers. For example, if the number was 27, Row 2, Column 7 was filled. Four such matrices were put on a page. A sample page of stimuli so constructed is shown in Fig. 1.

The stimulus on the upper left was designated the standard. The other three stimuli had the same amount of fill as the standard. The correct alternative among these three was constructed by replicating the standard and then moving the required number of randomly selected elements to new randomly selected positions that did not correspond to any of the filled squares in the standard. For example, if the stimuli are 30% filled, and we wish to accomplish a 30% distortion, nine squares are moved. The position of the correct alternative on the page was also determined randomly.

Incorrect alternatives were constructed in the same way as the standard, but were made to conform to several restrictions. The degree of correspondence between the standard and each incorrect alternative was held constant at the expected value for the given percentage of fill; four squares corresponded where the stimuli were 20% filled, nine where they were 30% filled, etc. A minor exception was the 10% fill case, where one square was made to correspond rather than the expected zero. The correspondence between the correct and incorrect alternatives was kept somewhat lower in order that the problems not be too difficult, but was again roughly proportional to the expected values. The values could not be kept arbitrarily low, since there were not enough empty squares to shift to in order to avoid both the standard and correct alternative in high fill cases. The values of correspondence chosen for the five fills were 0, 2, 6, 12, and 20 squares. Figure 1 is a case of 20% fill, 20% distortion, and close examination will show that it conforms to the above restrictions.

Four percentages of shift (called *distortion*) were used at each of the five levels of fill: 10, 20, 30, and 40%. Obviously distortions cannot reach a ceiling of 100%; the expected correspondence between a random alternative with 50% fill and a standard is 25 squares, which is the same as the correspondence between the standard and a "correct" alternative which is 50% distorted. Thus the maximal meaningful distortion for a 50% fill would be less than 50%.

Three different sets of four were constructed for each of the four levels of distortion and five levels of fill, giving a total of 60 problems.

Procedure. Six groups of 20 Air Force trainees at Lackland Air Force Base were used as *Ss*. Each *S* was given all of the problems. A presentation of one of each of the 20 types of problems constituted a "trial"; order of presentation of the 20 problems within each trial was randomized anew for every trial and every group. Each

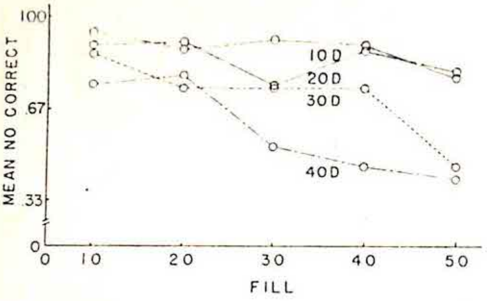


FIG. 2. Mean proportion correct at each fill level plotted for each distortion level.

group had three such trials, and thus saw each of the 60 stimuli once.

The *S*s were told that their problem was related to that of the radar navigator. The standard in the upper left corner was related to a briefing map, and the three alternatives were said to be like three parts of a radar scope display, one of which was the "target" represented on the briefing map. The *S* was to choose the correct target as quickly as possible. The *S* made his choice by writing the number of the correct alternative on his data sheet; the alternatives were numbered according to their position on the page. As soon as *S* made his choice, he wrote down the number of seconds that had elapsed since the problem had been presented; *E* counted elapsed seconds aloud, beginning 5 sec. after the problem was presented. A pilot experiment indicated that the time measure so obtained was satisfactory for the present purpose.

Stimuli were projected on a screen by means of a Baloptican projector. After instructions were completed and questions answered, *S*s were given five simple sample stimuli to make sure that the procedure was clear and fairly well practiced before beginning the actual data collection.

RESULTS

Figure 2 shows the average proportion of correct responses for the stimuli at each level of fill and distortion. There is the expected decrease in number of correct responses with increasing distortion. There is also a general tendency for correct responses to decrease with increasing fill, although this seems to be largely due to the marked decrease in correct responses at the highest distortion value for the 30%, 40%, and 50% fill cases. This large decrease also leads to a Fill × Distortion interaction. An analysis of variance showed that these three effects were significant at beyond the .01 level.

The results from the time scores agreed remarkably well with those from the number of correct responses. Comparison of Fig. 2 and 3 shows this close agreement. The correlations between the time and correct responses ranks in the F × D tables is .92. For time scores, Fill, Distortion, and the Fill × Distortion interaction were again significant at beyond the .01 level.

The effects of Trials, as trials were defined in this experiment, were significant ($P < .05$) but relatively slight. The effect on time scores was more stable than was the effect on number of correct responses. For time scores, trials remain significant when tested over their largest interaction with other variables, but this is not the case with the number of correct responses.

The Fill × Distortion interaction is perhaps the most interesting result. Attention was given to finding an index which would enable the difficulties of each of the classes of problems to be "predicted." It turned out that an "index of relative simi-

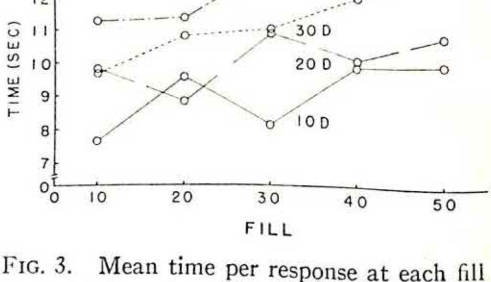


FIG. 3. Mean time per response at each fill level plotted for each distortion level.

larity" could account for a large amount of the variance within the Fill × Distortion table, whether for time or for correct response scores. The relative similarity was defined as the number of filled squares that corresponded between the standard and the correct alternative, minus the number of filled squares that corresponded between the standard and the incorrect alternatives, divided by the number of squares filled in each of the stimuli. In the 10% fill, 10% distortion, case the similarity is highest. Its value here is $\frac{9-1}{10} = .8$. Nine is the number of squares that are filled in the same positions on both the standard and the correct alternatives, one is the number filled in the same positions on both the standard and the incorrect alternatives, and 10 is the number of squares filled on each of the four matrices. Table 1 shows the values of the index for each cell of the Fill × Distortion table. All correct response scores for a given value of the index were totaled, and the mean number of correct responses obtained for each value of the index. A regression line was fitted to the resulting plot; the equation for the line of best fit was $C' = 138.85 + 280.63\ I$, where C' = the predicted number correct, and I the value of the index. The proportion of the variance accounted for by their correlation with the index was quite high, as indicated by the value of $r^2 = .88$.

The time scores were treated in the same way. The regression line was $T' = 5172.34 - 2912.76\ I$, where T' is the predicted mean time to respond. Again r^2 is quite high (.92). Although r^2 was high for both measures, it was obvious by inspection that an appropriate sigmoid curve would have fitted the points for number of correct responses even better than the straight line actually used. This is not true for the time scores.

TABLE 1

"RELATIVE SIMILARITY" FOR EACH VALUE OF FILL AND DISTORTION

Distortion	Fill				
	10%	20%	30%	40%	50%
10%	.8	.7	.6	.5	.4
20%	.7	.6	.5	.4	.3
30%	.6	.5	.4	.3	.2
40%	.5	.4	.3	.2	.1

In spite of the apparent success of the index, it was noted[3] that there were systematic differences in the time scores, due to distortion level, when the separate distortions were plotted against the values of the index; thus the index did not take sufficient account of distortion as it affected time. There were no residual effects of distortion on the correct response scores.

A revision was accordingly made in the index as applied to time scores; the revised index is simply $I + (.4 - D)$, where I is the original index and D the level of distortion. The new index incorporates distortion twice, once in the calculation of the original index and once in modifying it. When the time values were plotted against the revised index separately for each distortion level, there were no longer any systematic differences in the curves. It now seemed justifiable to combine the several levels of distortion according to their index values. The equation for the regression line fitted to the 11 values of the new index was $T' = 5150.79 - 2131.44\ I_1$, where I_1 is the new index. The value of r^2 increased to

---

[3] I am indebted to F. Attneave for a critical reading of this manuscript, and for pointing out these effects.

.94, despite the larger number of values and corresponding decrease in their reliability for the new index.

These regression lines are of course specific to the situation and probably hold only for this set of conditions. Some of the limiting factors are the fact that it is a three-alternative situation, there is a 20-sec. time limit for responding, and the stimuli were constructed according to the specified procedure. Nevertheless, the curves are at least a start toward the specification of difficulty of identifying stimuli, and they incorporate some of the factors that must be taken into account if stimulus difficulty is to be quantified.

DISCUSSION

The present study is at least a clear demonstration that any analysis and quantification of perceptual difficulty must consider parameters based on the relationship between test alternatives as well as those based on the stimulus that is initially studied. It must also consider the various relationships between the stimulus studied and the one designated correct. There is nothing wrong in relating the absolute characteristics of the stimulus to the difficulty in identifying it among a group of similar alternatives, but it seems that a complete picture of perceptual difficulty must consider what happens when the standard stimulus is distorted. For example, in the present study it matters little whether the grid is 10% or 50% filled if distortion is minimal, but it matters a great deal if distortion is maximal. Exclusive study of the no-distortion case might yield results which tend to mislead.

Another facet of the problem is the importance of the kind of operation used in producing distortion. French (1954b) added irrelevant elements and changed the orientation of the stimulus; other possible distortions might shift the outlines of solid figures, subtract elements, and the like. Different functions will no doubt be required to describe the difficulty following from each of the different operations. And each of these functions would probably be modified according to the importance of memory in the task.

It is possible that the task of describing perceptual difficulty might be reduced by referring the different classes of operations to a common mathematical basis, like that afforded by information theory. This is simple enough to do for the stimulus considered in isolation, but not so simple for the several kinds of distortion. For figures constructed on a model like the present one, the number of bits equals $\log_2 \frac{n!}{r!n - r!}$, where n is the total number of positions in the matrix, r is the number of filled squares, and $n - r$ is the number of empty squares. The function reaches its maximum at $r = n - r$, and is symmetrical about that point. For the present stimuli, the numbers of bits for 10% through 50% fill were approximately 44, 69, 85, 94, and 97. The use of the information model may or may not make it possible to simplify stimulus-response relationships for different types of stimuli.

SUMMARY

Stimuli were constructed in a 10 × 10 grid by a random sampling procedure. The number of elements filled in was varied in steps from 10% through 50%, and the number of elements rearranged from the standard stimulus to make the correct alternative was varied from 10% through 40%. The combination of fill and distortion resulted in 20 stimulus classes. The *Ss*' task was to choose the stimulus that most closely resembled the standard from the three alternatives presented. Each of 120 Lackland Air Force Base trainees had all 20 stimulus conditions. Fill, distortion, and their interaction all had significant effects on time to respond and on number of correct responses. An "index of relative similarity" was developed to "predict" problem difficulty, and showed a close relationship to both number of correct responses and time to respond. Some further suggestions were made for the quantification of perceptual difficulty.

REFERENCES

ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.

ATTNEAVE, F. Symmetry, information, and memory for patterns. *Amer. J. Psychol.*, 1955, **68**, 209–222.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N., & LEONARD, J. A. Stimulus correlates of visual pattern recognition. *J. exp. Psychol.*, 1956, **51**, 1–11.

FRENCH, R. S. The discrimination of dot patterns as a function of number and average separation of dots. *J. exp. Psychol.*, 1953, **46**, 1–9.

FRENCH, R. S. Identification of dot patterns from memory as a function of complexity. *J. exp. Psychol.*, 1954, **47**, 22–26. (a)

FRENCH, R. S. Pattern recognition in the presence of visual noise. *J. exp. Psychol.*, 1954, **47**, 27–31. (*b*)

(Received April 10, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

EFFECT OF AMOUNT OF PRECONDITIONING TRAINING UPON THE MAGNITUDE OF SENSORY PRECONDITIONING[1]

DONALD R. HOFFELD,[2] STEPHEN B. KENDALL, RICHARD F. THOMPSON,[3] AND W. J. BROGDEN

University of Wisconsin

The phenomenon of sensory preconditioning was presented by Brogden (1939) with the implicit assumption that its quantitative variation and the independent variables to which it is functionally related are similar to those of conditioning. Brogden (1950) reported briefly about two experiments on sensory preconditioning in which variation in the number of preconditioning training trials appeared to have no differential effect upon the magnitude of the phenomenon. Because of the fundamental relation of amount of training to amount of conditioning, it was assumed that there were errors of some kind in the experiments, rather than that sensory preconditioning differs in respect to this relationship from standard conditioning. Hoffeld, Thompson, and Brogden (1958) studied the stimuli time relations during preconditioning training for conditions of 0, .5, 1, 2, and 4 sec. precedence of tone over light and found maximum sensory preconditioning to tone for 4 sec. precedence. Since these results are not compatible with the results on time relations during standard conditioning, these authors suggest that sensory preconditioning may be markedly different from conditioning. It is possible that the variables known to affect conditioning may have an effect upon sensory preconditioning also, but that the nature of the effects will be different. Sensory preconditioning may be a phenomenon separate from standard conditioning, with a set of different relationships. Seidel (1959) in his recent review of studies of sensory preconditioning has also suggested this possibility.

The present study was planned with the assumption that sensory preconditioning is a phenomenon different from standard conditioning. Number of trials of preconditioning training was selected as the independent variable because of the fundamental character of the positive relation of amount of practice to amount of conditioning in learning fact and theory. A difference in this relationship for sensory preconditioning will be of primary importance in distinguishing it from standard conditioning. Three alternative hypotheses were considered: (*a*) that amount of sensory preconditioning will be a direct positive function of the number of trials of preconditioning training; (*b*) that amount of sensory preconditioning will be independent of number of trials of preconditioning training, once a sufficient number of trials has been given to produce the phenomenon; and (*c*) that as the number of preconditioning trials increases, sensory preconditioning will appear first at a minimum magnitude, then increase to a maximum, and thereafter decline.

---

[1] This research was supported in part by grants from the National Science Foundation and the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

[2] Now on active duty, United States Air Force, assigned to Human Engineering Laboratory, Rome Air Development Center.

[3] Now in Department of Medical Psychology, University of Oregon School of Medicine.

The initial experiment to test these hypotheses included values of 0, 10, 20, 40, 80, 200, and 400 preconditioning training trials.[4] The results show a significant amount of sensory preconditioning for all conditions greater than zero trials, but no significant differences in magnitude, although the amount of sensory preconditioning for 200 trials of preconditioning training was numerically greater than for any other condition. The experiment was expanded in two ways in order to provide differential support for either the second or third of the hypotheses presented above. One was the addition of the following values of preconditioning training trials: 1, 2, 4, 8, and 800. The other was an increase of *n* per group from 4 to 6. Thus, the second phase of the experiment involved the testing of 6 animals each for the 1, 2, 4, 8, and 800 trial groups and 2 animals each for the 0, 10, 20, 40, 80, 200, and 400 trial groups.[5] Further presentation will be as if there had been a single coordinated experiment from the beginning.

PROCEDURE

All experimentation was conducted in a sound-insulated, light-shielded room with the rotating cage of Brogden and Culler (1936) as the conditioning apparatus and under the general conditions described by Hoffeld, Thompson, and Brogden (1958). The conditions for the tone and light during preconditioning were the same as in the earlier study, except that the tone preceded the light by 4 sec., but overlapped for the 2 sec. duration of the light. These time relations were selected for preconditioning training because of the maximum amount of sensory preconditioning obtained for these conditions by Hoffeld, Thompson, and Brogden (1958). During the other phases of the procedure, avoidance conditioning to light, critical test with tone, and extinction to light, the duration of the light and the tone was 2 sec., and except for number of trials of preconditioning training, the procedure replicates that of Hoffeld, Thompson, and Brogden (1958).

The 72 cats of the experiment were assigned at random in two stages (see introduction) so that *N* for each of the 12 groups was 6. Three *E*s tested *S*s so that each *S* was tested by a single *E* for preconditioning training or for avoidance conditioning, critical test to tone, and extinction to light. In the majority of cases, a single *E* conducted both the preconditioning training and the subsequent phases of the experiment on the same *S*. However, *E* was not orthogonal in the experimental design. The daily experimental session involved 20 trials of the appropriate condition, except in the case of *S*s assigned to groups for 0, 1, 2, 4, 8, and 10 trials of preconditioning training. The preconditioning training of *S*s assigned to groups for 20, 40, 80, 200, 400, and 800 trials involved 20 trials per test period for 1, 2, 4, 10, 20, and 40 test periods, respectively. Following the preconditioning training, all *S*s received: (*a*) instrumental avoidance training with 2 sec. light CS and .1 sec. shock UCS until the conditioning criterion of 90% CR (18 or more cage-turning shock-avoiding CRs for the 20 trials) was reached; (*b*) tests for sensory preconditioning by presenting a 2 sec. tone CS until frequency of CR in a given test period was 10% or less; (*c*) extinction of the CR to the light CS until the criterion of 10% or less CR in a given test period was reached.

The *S*s were relatively quiet during preconditioning training. Since no cage-turning response occurred to any presentation of the preconditioning stimuli, this requirement of the sensory preconditioning paradigm (Brogden, 1939) was met. During instrumental avoidance training a CR was defined as a cage-turing response that occurred before the termination of the CS. The occurrence of a CR provided for avoidance of the electric shock UCS. This same criterion of response was maintained for the critical tests to tone and the extinction to light except that no shock was involved. The intertrial interval for all procedures varied between 15 and 30 sec.

---

[4] Consideration was given to the inclusion of control groups involving presentation of trials of tone alone during the pretraining period equal in number to the frequency of preconditioning training for the experimental groups. The failure in a number of studies of sensory preconditioning (Brogden, 1949; Chernikoff & Brogden, 1949; Coppock, 1958; Seidel, 1958; Silver & Meyer, 1954) to find any effect for noncontiguous presentation of the preconditioning stimuli makes the heavy investment of animals and time unnecessary.

[5] Hoffeld collaborated in the first and Kendall in the second phase of the experiment.

TABLE 1

MEASURES OF SENSORY PRECONDITIONING AND ACQUISITION

S	Group and Number of Preconditioning Trials											
	C 0	E_1 1	E_2 2	E_3 4	E_4 8	E_5 10	E_6 20	E_7 40	E_8 80	E_9 200	E_{10} 400	E_{11} 800
Frequency of Responses to Posttest Trials of Tone												
1	0	0	3	55	2	14	8	11	2	15	2	5
2	0	1	0	4	8	5	2	2	11	10	3	0
3	0	1	23	9	6	1	2	4	2	6	1	0
4	0	1	0	22	2	1	4	8	2	7	2	4
5	0	1	4	5	4	1	1	6	2	18	8	2
6	0	2	2	50	9	3	2	2	6	7	2	1
Mean	0.00	1.00	5.33	24.17	5.02	4.17	3.17	5.50	4.17	10.50	3.00	2.00
Test Periods to the Criterion of Conditioning												
Mean	7.83	8.50	6.83	11.67	5.50	5.83	7.33	6.17	6.33	4.63	6.50	9.50
Frequency of Responses to Extinction Trials of Light												
Mean	11.00	21.50	6.17	14.00	6.00	7.67	7.00	4.50	7.00	5.83	9.17	11.83

RESULTS

Examination of the data for acquisition, cross-modal stimulus generalization, and extinction (see Table 1) suggested the possibility of relationship between acquisition and each of the latter. Coefficients of correlation were computed for number of test periods to the conditioning criterion and both frequency of response and number of test periods for the tests to tone and extinction to light. No correlation coefficient is significantly different from zero. Analysis of variance of the acquisition data (see Table 1) produced an *F* ratio significant at the 5% level, but not at the 1% level. In view of the fact that the value of χ^2 is nearly equal to the value at which the hypothesis of homogeneity of variance must be rejected, the conclusion that there are significant differences in acquisition is tenuous. Since this result does not differentiate the control and experimental groups from each other, there is no evidence that acquisition influences the occurrences of sensory preconditioning as such. Differential magnitude of sensory preconditioning may be affected by acquisition, since Group E_3 shows the slowest acquisition and the maximum amount of sensory preconditioning. The results of Tukey's test for multiple comparisons (Ryan, 1959) of the acquisition data shows the mean for Group E_3 (largest mean) to be significantly different at the 5% level from the mean for Group E_9 (smallest mean). There are no other significant differences among these means.

Analysis of variance of the extinction data produced an *F* ratio whose value was considerably less than the value of *F* at the 5% level of significance. The consistent failure of extinction following the critical test for sensory preconditioning (Brogden, 1947; Chernikoff & Brogden, 1949; Hoffeld, Thompson, & Brogden, 1958) to demonstrate any secondary extinction effect between the experimental and control conditions suggests the conclusion that there is no such relation and that this procedure is completely ineffective for the measurement of sensory preconditioning.

A *t* test of the mean difference in frequency of response to the tone (cross-modal stimulus generalization) for the control group versus the experimental groups ($P = .0003$) established the reliability of sensory preconditioning for the experimental groups. An orthogonal polynomial analysis of variance (Grant, 1956) was completed on the measures of sensory preconditioning for all of the experimental groups and the results of this analysis are given in Table 2.[6]

TABLE 2

ANALYSIS OF VARIANCE OF RESPONSES TO TONE BY EXPERIMENTAL GROUPS

Source	*df*	*MS*	*F*
Between groups	9	274.30	4.03*
Linear	1	105.94	1.56
Quadratic	1	196.00	2.88
Cubic	1	217.70	3.20
Quartic	1	599.72	8.82*
Quintic	1	6.03	0.09
Sextic	1	571.21	8.40*
Residual	3	257.37	3.78
Within groups	50	68.03	

* $P = .01$.

Since the assumption of homogeneity of variance could not be accepted, the 1% level was used for the statistical tests of these data. The *F* ratio of the between and within variances is statistically significant, and

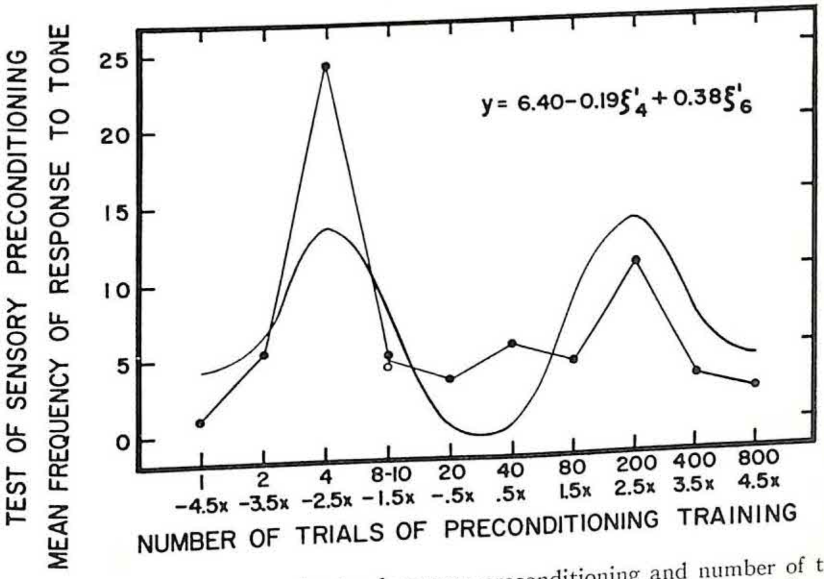


FIG. 1. Relation between magnitude of sensory preconditioning and number of trials of preconditioning training. The smooth curve is a plot of the orthogonal polynomial equation. The means of the data included in the orthogonal polynomial analysis of variance are represented by the closed circles. The open circle is the mean for Group E_5. The straight line curve connecting the means is drawn to the mean of the means for Groups E_4 and E_5.

---

[6] Since the orthogonal polynomial analysis requires equal intervals of the independent variable, a choice must be made between the inclusion of data for Groups E_4 and E_5. A toss of the coin eliminated Group E_5 from the data presented in Table 2 and the orthogonal polynomial equation presented in Fig. 1. An analysis that used the data for Group E_5 in the place of Group E_4 gave very similar results to those presented.

demonstrates significant differences between the experimental group means. The significance of the quartic and sextic sources of variation demonstrates that there are significant quartic and sextic components in the relation of magnitude of sensory preconditioning to number of preconditioning training trials. Figure 1 presents the best fitting orthogonal polynomial equation and the empirical data points. This equation is not an adequate representation of the relation for several reasons. It provides for a significant amount of sensory preconditioning (6.4 responses) at zero trials of preconditioning training and even greater magnitudes for negative trials of preconditioning training. It provides for no sensory preconditioning in the approximate range of 15 to 60 trials of preconditioning training. All of this is contrary to the empirically obtained results and must be rejected. In addition, the increase in amount of sensory preconditioning for trials in excess of 800 appears unlikely. On the positive side, however, the bimodal character of the function and trends of increase and decrease around each mode appear to be representative of the data. A further analysis of the data was made by means of Tukey's test for multiple comparisons (Ryan, 1959). The mean for Group E_3 (4 trials) is significantly different from all other means at the 1% level, except the mean for Group E_9 (200 trials) from which it differs significantly at the 5% level. There are no other significant differences between the means.

DISCUSSION

Of the three alternative hypotheses which the experiment was designed to test, the one relating amount of sensory preconditioning positively to number of trials of preconditioning training is not supported by the results and therefore is rejected. The results do not, however, provide clear and distinct support for either of the other hypotheses: that amount of sensory preconditioning is independent of number of preconditioning trials, once a sufficient number of trials has been given to produce the phenomenon, and that as the number of preconditioning trials is increased, the amount of sensory preconditioning will appear first at a minimum magnitude, then increase to a maximum, and thereafter decline. The results of the orthogonal polynomial analysis of variance and of Tukey's test for multiple comparisons are in opposition to the first of these hypotheses. There is the possibility, however, that the large amount of sensory preconditioning for Group E_3 (4 trials) is due to a slower acquisition of the CR. The results of the analysis of the acquisition data are indefinite and the failure to find any significant correlation between speed of acquisition and amount of sensory preconditioning gives no support to this view. Since sensory preconditioning occurs with one trial of preconditioning training, the modifying clause, once a sufficient number of trials has been given to produce the phenomenon, is superfluous. The remainder of the hypothesis, that amount of sensory preconditioning is independent of number of pretraining trials, in all probability should be rejected.

The remaining hypothesis has substantial support from the results of the experiment. There is a numerical increase in amount of sensory preconditioning from 1 to 2 to 4 trials of preconditioning training and there is a decline from 400 to 800 trials. The groups for 1, 2, and 800 trials are the only ones in which there are *S*s who gave no evidence of sensory preconditioning (1, 2, and 2 *S*s, respectively). The orthogonal polynomial analysis supports these trends. It provides, however, for no sensory preconditioning within the approximate range of 15 to 60 trials of preconditioning training. This is contrary to fact and cannot be accepted. The orthogonal polynomial analysis also provides two maxima, one at 4 trials and the other at 200 trials, whereas the results of Tukey's test provide for a maximum only at 4 trials.

If we consider that during preconditioning training the tone is a CS and the light a UCS (UCR implicit and unknown), the development of the CR (implicit and unknown) must be a function at least of the temporal relations of the light and tone (Hoffeld, Thompson, & Brogden, 1958) and of the frequency of presentation of these stimuli. The CR occurs after 1 trial and its strength increases up to 4 trials and thereafter declines, but not below the 1 trial level up to and including 800 trials. The decline may be a function not only of the number of trials but also of their distribution, since trials in excess of 20 were given by test periods of 20 trials. A decremental function due to repetition should show some recovery or less decline as a function of distribution. The massing of trials within a single test period with a standard intertrial interval should provide the most precise measurement of any decline in the magnitude of sensory preconditioning as a function of member of trials of preconditioning. The results of the present study support a hypothesis of an increase in magnitude of sensory preconditioning up to approximately 4 trials and a decline thereafter to 20 trials. The effect of trials beyond 20 is possibly contaminated by their distribution.

Any information about a CR established by the preconditioning trials of tone and light depends upon inference made from the results of test of cross-modal stimulus generalization. Since the phenomenon of sensory preconditioning is measured as a transfer effect, its measurement may be affected by the method of measurement. Thus any variable that affects stimulus generalization would affect the measurement of sensory preconditioning (Thompson, 1959). Use of the appropriate control, however, should demonstrate any effect of variables of stimulus generalization unless there are interactions of these variables with those affecting sensory preconditioning.

Measurement of cross-modal stimulus generalization by the extinction procedure involves the assumption that resistance to extinction is a direct measure of the strength of conditioning. Although the bulk of evidence supports this assumption, it is clear that extinction is not a highly reliable measure of the strength of a CR (Brogden, 1951). The unreliability of extinction measures undoubtedly contributes to the unreliability of the relation of number of preconditioning trials to magnitude of sensory preconditioning. In spite of this unreliability, however, the relatively large measure of sensory preconditioning for the 4-trial group is impressive. The mean of 24.17 responses to the tone represents a strength of CR between the tone and the light substantially greater than that obtained by 4 trials of standard conditioning procedure.

The results of the present study support the view that sensory preconditioning is a phenomenon of learning different from standard conditioning. Whether the further interpretation should be made that sensory preconditioning is a unique type or class of learning is another matter. Sensory preconditioning may occur concomitantly and be a part of standard conditioning. The procedures for standard conditioning are similar to those of preconditioning training and the phenomenon of sensory preconditioning may be necessary to the formation of standard conditioning. A primary difference between standard conditioning and sensory preconditioning is the presence in procedures for the former of a stimulus that evokes a clear and definite response and whose function can be spoken of in terms of reinforcement. Perhaps the only difference between standard conditioning and sensory preconditioning is in terms of performance and that the primary role of reinforcement is in terms of performance.

SUMMARY

The experiment was conducted on 72 cats who in groups of 6 were given 0, 1, 2, 4, 8, 10, 20, 40, 80, 200, 400, and 800 preconditioning training trials of a 6-sec. tone and 2-sec. light. After avoidance conditioning was established to the light, tests of cross-modal stimulus generalization to the tone were obtained as measures of sensory preconditioning. There was no response to the tone by the animals of the control group (0 trials) whereas all of the experimental groups showed significant amounts of sensory preconditioning. The magnitude of sensory preconditioning increases progressively through 1 and 2 preconditioning trials to a maximum at 4 and then declines to an approximately uniform level for 8, 10, 20, 40, and 80 trials, increases at 200 trials, and then declines progressively at 400 and 800 trials. An orthogonal polynomial analysis of variance shows significant differences between means and a trend with quartic and sextic components. Tukey's test for multiple comparisons indicates only that the mean for 4 trials is significantly greater than any of the other means. These results appear to establish a function of sensory preconditioning as trials increase from 1 to approximately 4, and then a decrease. The form of the function beyond the level of the decrease is in question.

Since the hypothesis that amount of sensory preconditioning increases as a function of number of preconditioning trials must be rejected, the results of the present experiment provide additional evidence that sensory preconditioning is a different kind of phenomenon from standard conditioning. Some of the theoretical considerations of this viewpoint are discussed.

REFERENCES

BROGDEN, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.

BROGDEN, W. J. Sensory pre-conditioning of human subjects. *J. exp. Psychol.*, 1947, **37**, 527–539.

BROGDEN, W. J. Sensory conditioning measured by the facilitation of auditory acuity. *J. exp. Psychol.*, 1950, **40**, 512–519.

BROGDEN, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

BROGDEN, W. J., & CULLER, E. Device for the motor conditioning of small animals. *Science*, 1936, **83**, 269–270.

CHERNIKOFF, R., & BROGDEN, W. J. The effect of instructions upon sensory preconditioning of human subjects. *J. exp. Psychol.*, 1949, **39**, 200–207.

COPPOCK, W. J. Pre-extinction in sensory preconditioning. *J. exp. Psychol.*, 1958, **55**, 213–219.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

HOFFELD, D. R., THOMPSON, R. F., & BROGDEN, W. J. Effect of stimuli time relations during preconditioning training upon the magnitude of sensory preconditioning. *J. exp. Psychol.*, 1958, **56**, 437–442.

RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56**, 26–47.

SEIDEL, R. J. An investigation of the mediation process in sensory preconditioning. *J. exp. Psychol.*, 1958, **56**, 220–225.

SEIDEL, R. J. A review of sensory preconditioning. *Psychol. Bull.*, 1959, **56**, 58–73.

SILVER, C. A., & MEYER, D. R. Temporal factors in sensory preconditioning. *J. comp. physiol. Psychol.*, 1954, **47**, 57–59.

THOMPSON, R. F. Effect of acquisition level upon the magnitude of stimulus generalization across sensory modality. *J. comp. physiol. Psychol.*, 1959, **52**, 183–190.

(Received April 22, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

SUPPLEMENTARY REPORT: THE RELATIONSHIP OF INDUCED MUSCULAR TENSION TO MANIFEST ANXIETY IN LEARNING[1]

O. IVAR LOVAAS

University of Washington

The present study investigated the possibility of reproducing the characteristic features of the performance of *S*s scoring high on the A scale as determined by Spence, Taylor, and Ketchel (1956), by inducing muscular tension experimentally in randomly selected *S*s. The learning task employed had been shown by Spence, Taylor, and Ketchel (1956) to discriminate between *S*s who score high and low on the A scale.

Method. The *S*s were 12 randomly selected, experimentally naive, male graduate students in psychology. A "tension" group (6 *S*s) learned a set of paired adjectives while pressing a Smedley hand dynamometer during the actual trials; a "no tension" group (6 *S*s) learned the set without any induced muscular tension.

A presentation device showed each stimulus word of a pair for 1.7 sec., then showed the response word for 2.3 sec., after which the stimulus word of the next pair immediately appeared. Inadequacies in the apparatus allowed an *SD* of about .1 sec. in the various exposure times. Intervals between trials were 10 sec.

A dynamometer was located on each side of the exposure window, about 25 in. apart and about 20 in. from *S*. The handles were padded. Tension was induced only in the preferred hand. The dynamometer pointer closed a circuit to light a bulb located 2 in. above one corner of the exposure window, whenever *S*'s pressure was ±3 lb. of the required pressure. During the learning trials each "tension" *S* squeezed this handle at half the maximal pressure exerted during the 30th sec. of a grip-pressure pretest. The tension signal light was kept on by *S* during each trial. The *S*s in the "no tension" group merely gripped the handle, and contact on the dynamometer allowed a pressure of only 3 lb. to turn the light off.

In order to avoid incentive motivation differences between the groups all *S*s were told that they would work for a predetermined number of trials, that if they got all correct one time they should attempt to get all correct again, and that they should do their best on each trial. A practice session of six trials on 10 paired nouns, with use of the dynamometers begun on Trial 3, was followed after 1 min. by a reading by *S* (aloud) of a nonpaired list of the adjectives to be learned. Errors of pronunciation were corrected. The set of adjectives was then learned by the anticipation method until *S* achieved two errorless trials or for a maximum of 33 trials. Pairs were shuffled after each presentation.

The set of adjectives used is presented by Spence, Taylor, and Ketchel (1956, Table 1, p. 307) along with the theoretical rationale and predictions for the performance of high- and low-drive *S*s on these pairs. Four out of the 10 pairs have stimulus and response words that are highly associated (high-associated pairs) and six pairs are such that the stimulus word tends to elicit more than one response word (competing pairs). Performance of high-drive *S*s should be inferior to low-drive *S*s on the competing pairs. On the four highly associated pairs, high-drive *S*s should show superior performance initially. With increased training, they should increasingly give response words *other* than those originally paired with the stimulus words. The initial superiority of the high-drive *S*s should disappear when these pairs become characterized by much response competition.

Results. Figure 1 shows the learning curves for the "tension" and "no tension" groups on the four highly associated pairs and on the six competing pairs. The insert in Fig. 1 is a rough reproduction of the curves from the study by Spence, Taylor, and Ketchel (1956). They used the same set of paired adjectives but employed *S*s scoring high and low on the Taylor Manifest Anxiety Scale. As can be observed, results from the two studies are strikingly similar. The results are consistent with the hypothesis that the performance of *S*s who score high on the A scale can be approximated by *S*s assumed to score lower on the A scale working under induced muscular tension.

Further support for the drive properties of induced muscular tension is to be found in the

---

[1] This article is taken from a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Washington, 1958. The author is indebted to Moncrieff H. Smith, Jr., under whose direction the thesis was prepared.

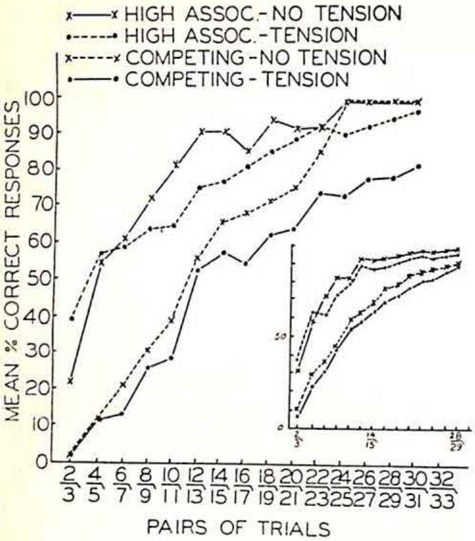


FIG. 1. Paired-associate learning as a function of induced muscular tension ("tension") and no induced tension ("no tension"). The insert shows the results of Spence, Taylor, and Ketchel (1956) employing the same paired associates, but with high-anxious (replacing "tension") and low-anxious (replacing "no tension") *S*s.

significant experimental Condition × Trials (2 through 10) interaction on the highly associated pairs ($F = 3.93$; $P < .01$). The interaction analysis was carried only through Trial 10 in order to preserve homogeneity of variances since many *S*s reached mastery beyond the trial and because previous studies (Spence, Taylor, & Ketchel, 1956; Spence, Farber, & McFann, 1956) had demonstrated that the reversal of superiority between high- and low-drive groups took place by Trial 5. On the Spence, Farber, and McFann assumption (1956, p. 299) that the highly associated pairs become characterized by an increase in response competition with increase in trials, the significant experimental Condition × Trials interaction supports induced muscular tension as a *D* variable.

REFERENCES

SPENCE, K. W., FARBER, I. E., & MCFANN, H. A. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

(Received January 31, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

SUPPLEMENTARY REPORT: RETENTION AND TRANSFER OF RESPONSES TO STIMULUS CLASSES

KENNETH E. LLOYD[1]

University of Virginia

Transfer to new instances of a class is less for *S*s who see only one instance of a class during original learning (L-I) than for *S*s who see more than one instance (Lloyd, 1959). The former *S*s (Specific Group) can learn a paired-associate task without recourse to a mediating response (the class name), and are not set to respond to a new instance at the transfer task (L-II). Once the mediating response is appropriate, however (i.e., in Class Groups when more than one instance is used at L-I), then transfer should be independent of the number of instances seen in L-I.

This study repeats the design of the former Exp. I with four changes: at L-I three groups (one specific and two class groups) saw one, two, or six instances; after the one-week recall the groups transferred to lists containing six, two, and six new instances, respectively; the 18 male *S*s in each group were volunteers from psychology classes at the University of Virginia; no latency measures were taken. At L-II the class groups should not differ, but should exceed the specific group.

Both the means and *F* ratios in Table 1 agree with those of the prior study. Tests of individual *df*'s indicate that at L-I the class groups required more trials than the specific group ($F = 14.05$; $P < .01$). Neither recall nor loss scores differed significantly. At L-II the class groups did not differ ($F < 1.00$), but they made more correct responses than the specific group ($F = 44.87$; $P < .01$). Two conclusions are appropriate: the recall results of the prior study are not limited to the number of instances that were used, and the variable contributing to the transfer results seems to be whether or not the class name is ever used as a mediating response.

---

[1] Now at Washington State College.

TABLE 1

Trials to Criterion (L-I), Recall, Loss Scores, and Correct Responses (L-II)

Group	L-I Trials		Recall		Loss Scores		L-II, Trial 1	
	Mean	*SD*	Mean	*SD*	Mean	σ_M	Mean	*SD*
1–6	6.11	1.59	20.50	3.49	2.99	.83	12.78*	4.38
2–2	8.56	2.93	19.11	3.63	3.76	.86	19.33	2.79
6–6	8.83	2.24	20.11	2.49	3.35	.64	19.22	2.25
F ratio ($df = 2,51$)	7.09*		<1.00		<1.00		22.46*	

* $P < .01$.

At L-II in Exp. I of the prior study the performance of Group ES improved rapidly over the first three subtrials. These *S*s could have been employing the class name, or could have been simply learning a second list. Group 1–6 should begin to employ the class name at L-II. Inspection of the first six subtrial means at L-II for these two groups reveals that Group 1–6 makes fewer correct responses at all points except the first subtrial. Since responding to classes of stimuli is more difficult, these data suggest that at L-II Group ES did not employ a class name, while Group 1–6 was learning to do so.

Reference

Lloyd, K. E. Retention of responses to stimulus classes and to specific stimuli. *J. exp. Psychol.*, 1960, **59**, 54–59.

(Received June 1, 1959)

Journal of Experimental Psychology
Vol. 59, No. 3, 1960

SUPPLEMENTARY REPORT: TWO-CHOICE BEHAVIOR OF MONKEYS[1]

WILLIAM A. WILSON, JR.

University of Colorado

Monkeys trained in a .75–.25 noncontingent probability learning situation quickly reached an asymptote of approximately 100% choice of the .75 side, when correction was not permitted (Wilson & Rollin, 1959). The data suggested that *S*s would have reached the same asymptote, albeit more slowly, when immediate correction was permitted, but the experimental sophistication and sequence of training of *S*s did not permit a definite conclusion on this point. This paper reports on the behavior of naive *S*s under the same correction procedure, as well as on *S*s subjected to a "guidance" procedure in a contingent situation (Bitterman, Wodinsky, & Candland, 1958).

Method. Eight immature naive cynomolgus monkeys were used. A two-choice apparatus was used, modified slightly from one previously described. Two black plastic boxes, which served as covers for shallow food wells, were hinged at the back and weighted so that if the front of a box were slightly raised the box would open completely. A clip could be attached to connect the boxes so that opening either would cause both to open. Two "half-clips" could be attached to present a similar appearance, although leaving the boxes independent.

The *S*s were presented with a .75–.25 random situation, as described in the previous report; for half the *S*s in each group the left was the more often baited side. For the Correction Group (four *S*s), the clip was used, and, as in Part 3 of the earlier study, after *S* opened an unbaited box, he was permitted to retrieve the peanut from the other food well. For the Guidance Group, the half-clips were used. After *S* opened an unbaited box, the screen was closed, and then reopened, allowing *S* now to open the only closed box, the baited one.

Results. On the first day of testing, the two groups had similar wide ranges (Correction: 21.9%–100%; Guidance: 28.1%–84.4%) and means of approximately 60% choice of

---

[1] This research was done at the Institute of Living and was supported in part by the Research and Development Division, Office of the Surgeon General, Department of the Army, under Contract No. DA-49-007-MD-763. Most of the data were gathered by Ann Warren.

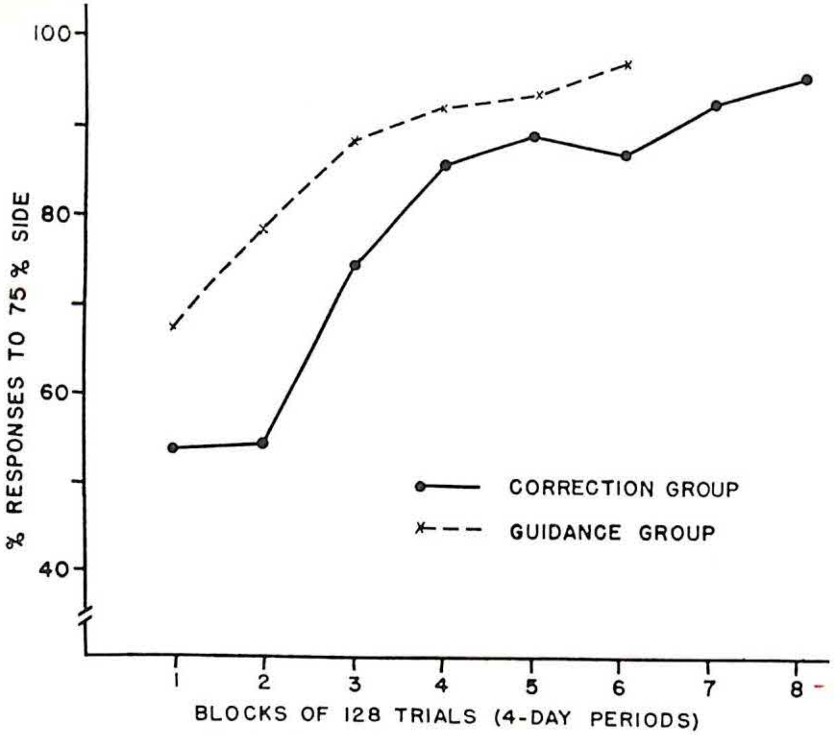


FIG. 1. Mean performance of the two groups.

the more often baited side. The course of mean performance is shown in Fig. 1. Despite the consistent difference in means, there was widespread overlap between the two groups at all points represented. On the last block of trials diagrammed, each of the *S*s made over 95% choices of the .75 side.[2]

Less ambiguous evidence is thus presented for the conclusion that monkeys will approximate 100% choice of the more often baited side in a .75–.25 noncontingent situation, even when correction is permitted. Furthermore, the same response pattern is seen when a guidance method is used in a contingent situation—although a similar method leads to event-matching behavior in fish (Bitterman et al., 1958).

REFERENCES

BITTERMAN, M. E., WODINSKY, J., & CANDLAND, D. K. Some comparative psychology. *Amer. J. Psychol.*, 1958, 71, 94–110.

WILSON, W. A., Jr., & ROLLIN, A. R. Two-choice behavior of rhesus monkeys in a noncontingent situation. *J. exp. Psychol.*, 1959, 58, 174–180.

(Received April 27, 1959)

---

[2] One additional *S* was tested under the procedures of the Guidance Group, but using a regular RRRL pattern for the placement of reward rather than a random .75–.25 schedule. The behavior of this *S* was not obviously affected by this difference in procedure; his percentages of choice of the right side on successive 128-trial blocks were 67, 70, 73, 84, 90, and 100.

Journal of
Experimental Psychology

VOL. 59, NO. 4 APRIL 1960

PERFORMANCE CHANGES IN EYELID CONDITIONING AS RELATED TO THE MOTIVATIONAL AND REINFORCING PROPERTIES OF THE UCS[1]

M. A. TRAPOLD AND K. W. SPENCE

State University of Iowa

In a recent series of lectures (Spence, 1956) the possibility was suggested that classical aversive conditioning is governed by a reinforcement principle. According to this view the development of the associative factor, habit strength (H), of a classically conditioned defense response is a function of the occurrence and intensity of the reinforcing UCS. This assumption, taken in conjunction with the existing assumption that drive level (D) is also a function of the intensity of the UCS, implies that the differences in performance level obtained with different intensities of the UCS reflect differences in both H and D. The problem of testing this theory thus requires that some means be found to separate out these two hypothetical functions of the UCS.

Since these lectures, several experimental studies, in which it is believed this separation was achieved, have reported findings that support this reinforcement interpretation of classical aversive conditioning (Spence, Haggard, & Ross, 1958a, 1958b). These experiments have involved holding the drive effects of the UCS constant while varying its reinforcing effects. In order to understand the logic of these experiments, as well as the one to be presented here, it is necessary to consider the source of motivation or drive in classical aversive conditioning. Since a CR occurs before the application of the UCS, we cannot appeal to the UCS on that particular trial as the source of drive. Furthermore, it has been shown (Spence & Runquist, 1958) that the drive cannot arise from emotional (fear) responses conditioned to the CS, since the latency of such fear responses is much longer than the 300–500-msec. latency of the conditioned eyeblink. Hence, the assumption has been that the drive level (D) is determined by previous presentations of the UCS. More specifically, the UCS is assumed to arouse a hypothetical emotional response (r_e), the effects of which are assumed to persist at least until the following trial. There is also the possibility that these emotional responses may become conditioned to other cues of the general stimulus situation. In any event, the assumption is made that the level of this hypothetical emotional activity is fairly constant from trial to trial, and that this level is, in large part, determined by the strength of the UCS. Drive level (D)

---

[1] This study was carried out as part of a project concerned with the influence of motivation on performance in learning under Contract N9 onr-93802, Project NR 154–107 between the State University of Iowa and the Office of Naval Research. Acknowledgement is made to John Hunter and Robert Weyant who served as research assistants in the above investigation.

is, in turn, assumed to be directly related to this level of emotional activity.

One of the implications of this analysis of the source of motivation in eyelid conditioning is that any application of a noxious stimulus in the situation will increase the level of r_e, and therefore the level of D. The above mentioned experiments capitalized on this implication by providing a technique by which D could be held constant for two groups of *S*s, and yet allowed the strength of the reinforcing UCS to be varied. Briefly, the technique consisted of using two strengths of air puff with each *S*. In the high reinforcement condition, the stronger air puff was always paired with the CS as a typical UCS, while the weak air puff was presented equally often in such a way that conditioning on these trials was precluded. In the low reinforcement group the reverse conditions held, the weak puff being paired with the CS and the strong puff always being presented so as to preclude learning. This procedure thus equated the average puff strength for both groups, and hence their drive level, but allowed the strength of the reinforcing UCS to be varied. Making use of the basic equation that performance (R) is proportional to $E = H \times D$, then, to the extent that D is equated for the two groups by this technique, any differences in their performance must be attributed to differences in H.

The results of these experiments have shown that when drive is so equated the asymptotic level of performance varies directly with the strength of the reinforcing UCS. These results have been interpreted as supporting the notion that the growth of H is a function of the strength of the reinforcing stimulus, and hence that a reinforcement principle operates in classical aversive conditioning. These experiments have also confirmed the assumption that D is a function of the intensity of the noxious stimulus employed, and that the unpaired puff is effective in manipulating the level of drive. Thus a group conditioned with a weak puff paired with the CS and a strong puff presented alone performed at a higher level than a group receiving a weak puff on both paired and unpaired trials.

One of the properties that has been ascribed to habit strength is that once accrued, habit strength is permanent, and cannot directly be reduced. In the 1943 version of his theory, Hull explicitly assumed that habit strength was a relatively permanent organization left within the nervous system of the organism. This assumption is also implicit in a more recent formulation (Hull, 1951). If this assumption about habit strength is taken in conjunction with the analysis of the role of the UCS presented in the above paragraphs, some interesting predictions concerning the effects upon performance of shifting the strength of the reinforcing UCS become evident.

(*a*) For a group, conditioned to the limit of learning with a weak air puff paired with the CS and a strong puff presented unpaired, and then switched to the reverse conditions, i.e., the strong puff paired and the weak puff unpaired, performance should show an increase following the shift, eventually reaching the level of a group trained under the high reinforcement condition from the beginning of training. The increase in the strength of the reinforcing UCS raises the level to which habit can grow, and the increase of habit to this new level is reflected in a performance increase.

(*b*) If a group, conditioned to the limit of learning with the strong puff paired with the CS and the weak puff presented alone is switched to the condition where the weak puff is now paired and the strong unpaired, no performance changes should be exhibited following the shift. Since habit strength is assumed not to decrease once accrued, and since the same drive level is maintained, there should be no change in performance.

(*c*) If a group, conditioned to the limit of learning with the strong puff as the reinforcing UCS, and the weak puff unpaired, is then switched to the condition in which both the paired and the unpaired puff are weak, we would expect a decrease in performance. This implication follows from the analysis of the source of motivation in aversive conditioning. In this instance the switch decreases the average puff strength, which would lead to a decrease in the level of hypothetical emotional activity and hence the level of *D*. To the extent that Prediction *b* is confirmed, performance changes in the present shift would reflect only a decrease in drive level. However, if the switch outlined in Prediction *b* does lead to a performance decrease, then the present group would reflect both a decrease in drive and a decrease in the reinforcing UCS.

The experiment reported here was conducted to provide data on these three predictions.

METHOD

Subjects. A total of 64 men from an introductory course in psychology was employed. Of these, five were eliminated for having met the criterion previously determined as defining a voluntary responder (Spence, 1953). The data of four other *S*s were also discarded, two for giving CRs on the pretest trials, and two who blinked continuously, making scoring of their records impossible. The remaining 55 *S*s provided three groups, one of which had 15, and two of which contained 20 *S*s.

Apparatus and method of recording. The *S* was seated in a dental chair in a semidarkened soundproof room. The *E* was in an adjoining room in which the stimulus controls and recording equipment were located. The apparatus for presenting the stimuli and recording the eyeblinks was identical with that used in previous studies from this laboratory (cf. Spence, 1953).

The CS consisted of the increase in brightness of a 6-cm. circular milkglass disk from a level of less than .004 mL. to 2.95 mL. The duration of the CS was 550 msec., with the UCS occurring 500 msec. after the onset of the CS. The duration of the UCS, an air puff of either .33 or 2.0 psi, applied to the right eye through a .062 in. diameter orifice, was limited to 50 msec. by a 100-v., 60-cycle ac-operated solenoid valve controlled by an electronic timer.

The word "ready," delivered by *E* over an intercom system, preceded each trial by 2, 3, or 4 sec., according to a prearranged schedule. The *S*s were instructed to blink upon presentation of the ready signal, and then fixate the circular disk in front of them. A CR was recorded whenever the record showed a deflection of 1 mm. or more in the range 200–500 msec. following the onset of the CS. Responses in the interval 0–200 msec., which were infrequent, were classified as original responses and not included in the data.

Conditioning procedure. Following the reading of the instructions, each *S* was given three presentations of the CS alone followed by a single presentation of the 2.0-lb. UCS alone. Immediately following these preliminary trials, conditioning was begun. Each *S* received a total of 260 trials, 130 conditioning trials which involved the paired presentation of the CS and UCS, and 130 which involved only the UCS following the ready signal. The order of presentation of these paired and unpaired trials was prearranged according to an irregular order in which the number of each was equalized in blocks of four trials. Intertrial intervals of 15, 20, and 25 sec., averaging 20 sec. and arranged according to a fixed schedule, were used throughout the experiment. Immediately following Trials 80 and 160, the experiment was interrupted while *E* checked the adjustment of the apparatus. No additional lights were turned on in the experimental room during this check, and it rarely took longer than one minute.

Experimental design. The experiment consisted of two phases. In Phase I, which lasted for 180 trials (90 conditioning trials), Groups A and B ($N = 20$ each) received identical conditions. For these groups the CS was always paired with the strong puff, the weak puff always being presented alone. For Group C ($N = 15$) the reverse conditions held, the weak puff always being paired with the CS and the strong puff always being presented alone.

Immediately following Trial 180, with no break in the continuity of the experiment, Phase II was begun. In this phase, Group A was switched to the reverse of its conditions in Phase I, the weak puff now being paired with the CS, and the strong puff being presented alone. Similarly, Group C was also switched to the reverse of its Phase I conditions, with the strong puff now paired with the CS and the weak puff unpaired. Group B was switched to receiving the weak puff on both the paired and the unpaired trials.

RESULTS

Figure 1 presents the conditioning curves for the three groups. Here the

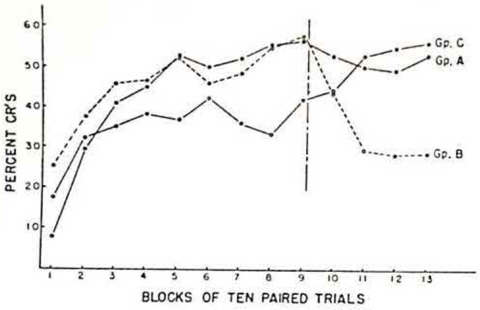


FIG. 1. Acquisition curves showing the percentage of CRs in successive blocks of 10 paired conditioning trials. The vertical line shows the point at which the shift was made.

percentage frequency of CRs is plotted against the number of conditioning trials (paired CS and UCS), which, of course, were the only trials on which CRs could be given. The vertical line represents the point at which conditions were switched from Phase I to Phase II. Groups A and B, which had identical treatment up to this switch point, can be seen to have attained almost identical performance levels. In contrast, Group C, which had the weak puff as the reinforcing UCS, reached an appreciably lower performance level than those of Groups A and B. Evaluation of this difference over the last 20 preshift conditioning trials revealed that the difference is statistically reliable ($P = .026$).[2]

Following the shift at Trial 90, it can be seen that Group B dropped quite rapidly down to a new level of performance. Comparing the last 20 trials before the shift (71–90) with the last 20 trials after the shift (111–130), reveals that this decrease is highly reliable ($P < .01$). Conversely, Group C increased following the shift, and came to approximate the preshift performance levels of Groups A and B. Again comparison of the last 20 preshift trials with the last 20 postshift trials reveals that this increase is significant beyond the 1% level. In contrast to Groups B and C, Group A maintained a relatively stable level of performance following the shift. Although Group A did show a slight decrease over Trials 90–120, the maximum amount of this decrease was only 7%, which is easily attributable to chance ($P = .29$).

DISCUSSION

The results of the preshift performances replicate those of earlier experiments in showing that when precautions are taken to insure equalization of drive level (*D*), performance is an increasing function of the strength of the reinforcing UCS. This is in line with the reinforcement position with respect to classical defense conditioning as outlined in the introduction. The postshift results also clearly support the predictions made by the reinforcement interpretation of this type of learning. When the strength of the reinforcing UCS was increased for Group C, hence raising the level to which *H* could grow, this group showed a corresponding increase in performance to the level of the previous high-reinforcement groups. The effect of the shift in strength of the reinforcing UCS on Group A, however, was quite different. In accord with prediction, habit strength being permanent, the decrease in the reinforcing UCS had little or no effect on performance. Taken together the results of these two groups lend considerable support to the notion that habit strength is an increasing function of the strength of the reinforcing UCS, and to the assumption that habit strength is a permanent acquisition that does not decrease when conditions are changed in the direction of being less favorable for habit *growth*.

---

[2] Because the distribution for frequency of response deviated considerably from normality, Mann-Whitney *U* tests were used to evaluate all between-group comparisons. Distributions of difference scores, however, were approximately normal, and hence *t* tests for related measures were used in all within-group comparisons. All reported probabilities are for tests of a two-tailed hypothesis.

Taken in conjunction with the post-shift results of Group A, the decrease in performance following the shift for Group B can be said to represent the effects of decreased drive only. Since Group A showed no effect from merely decreasing the reinforcing UCS, the drop in Group B's performance must be due to the lower over-all puff strength, which, according to this analysis, is a primary determinant of drive level.

SUMMARY

This study was concerned with several implications of a theory that assumes habit strength (H) to be a function of the intensity of the UCS in classical aversive conditioning. Groups A and B were conditioned with a strong air puff as the reinforcing UCS, a weak puff being presented equally often unpaired with the CS. Group A was then switched to the weak puff as the paired UCS and the strong puff unpaired. Group B was switched to the weak puff on both the paired and unpaired trials. Group C was conditioned with the weak puff paired with the CS and the strong puff unpaired, and then was switched to the reverse condition, strong puff paired and weak puff unpaired.

The preshift performances of Groups A and B were significantly higher than the performance of Group C. Following the shift, Group C gradually increased in performance to the preshift level of Groups A and B. Group B decreased significantly, and Group A showed no change in performance.

These results were interpreted as indicating that H is a function of the reinforcing UCS, that the extra unpaired puffs are effective in manipulating D, and that once accrued, H is permanent and does not decrease when conditions less favorable to habit *growth* are introduced.

REFERENCES

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **42**, 57–63.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411. (a)

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. Intrasubject conditioning as a function of the intensity of the unconditioned stimulus. *Science*, 1958, **138**, 774–775. (b)

SPENCE, K. W., & RUNQUIST, W. N. Temporal effects of conditioned fear on the eyelid reflex. *J. exp. Psychol.*, 1958, **55**, 613–616.

(Received June 1, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

ABSOLUTE JUDGMENTS OF ODOR QUALITY[1]

TRYGG ENGEN AND CARL PFAFFMANN

Brown University

This paper is concerned with the recognition of odor qualities and the application of information theory to the evaluation of the human observer in this modality. Garner and Hake (1951) have shown that an information measure, the bit, gives an index of the maximum number of single stimulus categories that can be discriminated without error by the method of absolute judgment. The present experiments were designed to determine the greatest number of different qualities that could be identified correctly by a label with different kinds, intensities, and numbers of odorants.

METHOD

Judgments were obtained from five *S*s, one man and four women. All were college students who were paid $1.00 per hour. They had served previously in olfactory experiments on intensity differences. The experiments reported here extended over a period of 12 months and were interspersed among several other experiments on threshold and scaling.

The experiments were performed in a well-ventilated and air-conditioned 5 × 7-ft. room at a temperature approximately 24°C. In the first part of the experiment *S* learned to identify the stimuli. He was asked to label each odorant with his own association to that odorant (avoiding the restriction of an arbitrary code). This was repeated for each sample of odorants used. After the labels had been chosen, they were checked a second time to eliminate duplications and to give *S* the opportunity to make changes.

The odorants were kept out of sight behind a shield. Only one odorant was presented at a time in random sequences, and *S* was given unlimited time to identify it with the previously assigned label. Each odorant in a set was presented 24 times for a total of 70 judgments of 5 odorants to 1152 judgments for 48 odorants. A correction procedure was used such that the correct label was indicated by *E* after *S* made his judgment. After a period of about 12 min., *S* was given a rest period while one or two other *S*s were run. This arrangement seemed optimum for control of fatigue and adaptation. It should be noted that with five *S*s it was not possible to obtain a balanced presentation of conditions, but no evidence was obtained that this restriction on experimental design had any appreciable effect on the results.

Two sets of 36 odorants were used. They are listed in Table 1 as different and similar. The set of so-called different odorants was selected to sample a wide variety of odorants, e.g., fragrant, spicy, acid, burnt, etc., and the similar set was selected to sample only sweet-fruity odorants. The latter set was obtained largely from Givaudan-Delawanna, Inc. and were described in their catalogue (No. 12, 1952) as having sweet and fruity odors. The remaining odorants were obtained from Fritzsche Brothers, Inc. and Brown University chemistry stock room. A few odorants (e.g., phenylethyl alcohol) appear in both sets. When fewer than 36 odorants were used in an experiment, they were selected at random from among the 36 odorants. In Table 1 a "5" opposite the odorant means that this odorant was included in the sample of 5 odorants, "7" that it was included in a sample of 7, etc. A "16" opposite the odorant indicates that the odorant was one of the 16 odorants presented at three levels of intensity. When the experimental procedure required it, geometric concentration dilution series were prepared of the odorant with a pipette technique, using benzyl benzoate as the dilutent (Beck, Kruger, & Calabresi, 1954). The odorants were presented in 10 × 75-mm. Pyrex test tubes stopped with an aluminum-wrapped cork.

Experiment I was concerned with the effect of variations in intensity and type of odorant on the amount of information transmitted. In a preliminary experiment each member of an undergraduate class of 40 students had been presented one test tube from a dilution

---

[1] This research was supported by the General Foods Corporation. Part of the results was presented at the Thirtieth Annual Meeting of the Eastern Psychological Association in Atlantic City, April 4, 1959.

TABLE 1

Chemical Agents Used in the Experiments
(Numbers Opposite Agents Are Explained in the Text)

YM

Different	Similar
Acetanisol—16	Acetate C-8—24, 10
Acetic acid—24, 15, 10, 7, 5	Acetate C–9—24, 10, 7, 5
Acetone—24, 15	Acetate C–10—24, 7, 15
Amyl acetate—24, 16, 15	Acetate C–11—24, 10
Allyl Caproate—24, 16, 15, 10, 7	Aldehyde C–7
Benzaldehyde, N.F.—16	Aldehyde C–8
Benzene—24, 16, 10, 7	Aldehyde C–10—24, 10
Benzyl Cinnamate—7, 5	Aldehyde C-14 (peach)—24, 15
N. Butanol—24, 7	Aldehyde C–14 (pure)—24, 10
N. Butyric acid—10	Aldehyde C–16—24, 15
Camphor (lt. syn.)—24, 16	Aldehyde C–18—24, 10
N. Caprylic	Allyl Caproate—24, 15, 7
Citral—24, 15	Amyl acetate—24
Clove oil—24, 15	Amyl Phenylacetate—24, 15, 10, 5
Coumarin—24, 15	Amyl Propionate
Diacetone alcohol—24	Benzyl Acetate Coeur—10
Ethyl acetate—7	Benzyl Butyrate—24, 15, 10
Eugenol—24, 16, 15, 10, 7	Cinnamyl acetate—24, 15, 10, 5
Guaiacol—24, 16, 15	Cinnamyl Butyrate—24, 15
Heptanol—24, 16, 15	Cinnamyl Propionate—24
Heptane—24, 15, 7	Citral—24, 10
Hexane, N.—24, 16, 10	Citronellyl acetate—24, 5
Indol—24, 15, 10	Citronellyl Butyrate—24, 15, 7
Isopropanol—16, 15, 10, 5	Dimethyl Anthranilate
Linalool—24	Geranyl Butyrate
Menthol—24, 16, 10	Geranyl Propionate
Methanol—16, 10, 5	Linalool—24, 15, 7
Methyl Salicylate—24, 16, 5	Menthol
Musk—24, 16	Methyl Anthranilate Standard
Novoviol, alpha—15	Methyl Salicylate—24, 15, 7
Phenylethyl alcohol—24, 16, 15	Phenylethyl acetate—24, 15, 5
Pyridine—24	Phenylethyl alcohol—24, 15
Safrole—24	Phenylethyl Isovalerate—24, 15, 10, 7
Vanillin—24	Raspbery aldehyde
Butyl acetate—15	Tolyl acetate—24, 15
Nitrobenzene	Vanillin

series of amyl acetate and was asked whether the odorant was "weak," "medium," or "strong." It was learned that a 6.25% solution of amyl acetate was, on the average, judged to be of medium intensity or "strength." This standard solution was then cross-matched by the five *S*s used in the present experiments with the 23 other different odorants in Table 1. Ten such cross-matches were made by each *S* for each odorant, varying the order and range of concentration dilutions of the comparison odorant to prevent context effects. This experiment yielded 24 odorants for each *S* of subjectively equal and medium intensity but different in quality for the first series of judgments. The same 24 odorants, but each in a 100% solution, were presented for a second series of judgments. A third selection of odorants was made of 24 odorants of similar odor qualities, each being presented in 100% solutions, as in the second series above.

Experiment II evaluated the information transmitted when both qualities and intensities of each odorant varied. Sixteen different odorants were selected, each in three concentration dilution values, 100, 25, and 6.25%. All of these concentration dilution values were clearly above threshold and more than 1 jnd apart. An earlier experiment (Engen & Pfaffmann, 1959) had indicated that these *S*s could transmit at least 1.5 bits of information about odor intensity. The purpose of this experiment was to determine whether or not the information transmitted for intensity would combine with the information transmitted for quality.

Experiment III was concerned with "maxi-

TABLE 2

Means and *SD*s of Bits of Information Transmitted by 5 *S*s for Various Samples of Odorants

Sample of Odorant	Mean	*SD*
Different odorants of equal and medium intensity	4.00	.17
Different odorants unequal intensities	4.03	.19
Similar odorants unequal intensities	3.86	.43

mum channel capacity" (Alluisi, 1957) for odor quality and the extent to which the amount of information transmitted in the experiment above may be limited by the number of stimuli presented for judgment. Thus 5, 7, 10, 15, 24 and 36 odorants in 100% solutions were presented consecutively to each *S* under the same conditions as above. Both similar and different odorants were used.

Results

The amount of information transmitted by each *S* under each condition in Exp. I was determined by a method suggested by Garner and Hake (1951). The average amount of information for all *S*s is presented in Table 2. First, the results indicate that the average *S* could transmit about 4.00 bits of information about odor stimuli which represent a wide variety of qualities (i.e., different odorants). According to information theory, then, *S* could use about 16 categories (the antilog of 4.00) without error in identifying odors. Second, there is the suggestion that this result will not vary appreciably with the intensity of these odorants, i.e., the distinctiveness of the odorants did not seem to be impaired by matching them in subjective intensity, because the matched odorants of medium intensity yielded about the same result (4.00 bits) as when they were unmatched (4.03 bits). Third, the selection of qualities to be judged is an important factor, because when the odorants were all similar (sweet-fruity) only 3.86 bits of information (14 categories) were transmitted. The similar odorants also seem to yield a more variable performance, as indicated by *SD*s which give an index of individual differences.

It should be noted that results for unmatched intensities must be interpreted with caution, because the two variable aspects of the stimulus, quantity and intensity, are not orthogonal. The presentation of each odorant at 100% concentration dilution provides chemically specified qualities but (psychological) intensities of indeterminate values, thereby precluding equally frequent occurrences of given quality and intensity combinations as well as a more appropriate and complex analysis (McGill, 1954) of the data. This is a statistical problem separate from the problem of psychological additivity of quality and intensity.

Quality and intensity were orthogonal by design in Exp. II in which each of 16 odorants were presented at three levels of intensity. These multidimensional judgments were analysed according to McGill's multivariate procedure, and the result of this analysis is presented in Table 3. The table shows that 3.51 bits of

TABLE 3

Means and *SD*s of Bits of Information Transmitted by Five *S*s for 3 Levels of Intensity for 16 Odorants

Transmission	Mean	*SD*
Quality—Responses	3.51	.14
Intensity—Responses	.27	.05
Interaction	.09	.01

information were transmitted about quality and .27 bits about intensity. Taking into account the small interaction of .09 bits per stimulus, it is indicated that 3.60 bits were transmitted about quality for a specified intensity and .36 bits about intensity for a specified quality. It is evident again that intensity did not add much to the distinctiveness of the odor stimuli.

Figure 1, which presents the data from Exp. III in the dashed-line curve, shows that the data presented in Table 2 are not limited by the 24 categories presented for judgment. Performance improves up to 24 categories but no further improvement results from presenting a larger number of categories, while efficiency is greatly reduced. Moreover, Fig. 1 shows that the difference between different and similar odorants is verified. It is interesting to note that performance was never perfect even when fewer than 16 alternative stimuli were presented. According to these results there is a bias in the prediction of the number of categories *S* can use without or with a minimum of error. It does not, however, affect seriously the interpretations made of the results.

DISCUSSION

It is often assumed that the human *S* has a great capacity for discriminating and remembering odor qualities (Behnke, 1954). There has also been much discussion about the difference in the human *S*'s ability to remember and discriminate odor intensity and odor quality. Information theory provides one measure relevant to this problem since the bit permits the comparison of absolute judgments obtained under different experimental situations. Miller (1956) has already done this for several sense modalities.

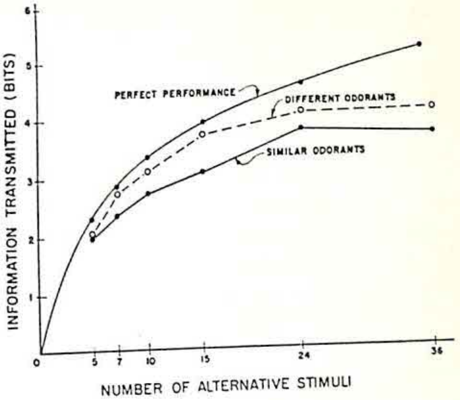


FIG. 1. Channel capacity for absolute judgments of odor qualities. Amount of information transmitted against number of alternative stimuli for two kinds of odors (described in text).

While a relatively practiced *S* can transmit a maximum of 2.0 bits of information about odor intensity (Engen & Pfaffmann, 1959), the present results indicate that his channel capacity is twice as high for odor quality. Both sets of results appear to have some generality since they were obtained under a reasonably wide range of experimental conditions. As to the comparison of absolute judgments of odor intensity and quality, then, the informational analysis of absolute judgments probably indicates that the popular notion is in error. Given some practice, the human *S*'s ability to judge odor intensity approaches his ability to judge auditory and visual stimuli. His ability to identify odor quality is not phenomenal but falls within the range of his ability to identify multidimensional stimuli in other sense modalities (Miller, 1956).

One might raise the problem of the generality of the present results to other odor samples. It is practically impossible to sample all the odorants that could have been used in these experiments. However, while the present quality experiments indicate a significant variation in amount of transmitted information due to the selection of odorants (similar and different), this variation is not large enough to change drastically the span of absolute judgments. For similar odorants the results indicated 3.86 bits and for different odorants, 4.03 bits. In this connection it is interesting to note that as the number of categories presented for judgment was increased for different odorants (Fig. 1) even very familiar odors, e.g., vanillin, were confused with others. When similar odorants were presented, this confusion of vanillin occurred in still fewer categories. The experience derived from the present experiment suggests that at best only an extremely careful selection of odorants would raise the channel capacity for quality under the present conditions.

The results also have some further bearing on the problem of multidimensional judgments. It was surprising, judging on the basis of similar experiments in other sense modalities (Beebe-Center, Rogers, & O'Connell, 1955; Eriksen, 1954; Klemmer & Frick, 1953; Pollack, 1953; Pollack & Ficks, 1954), that the presentation of three intensity levels of each of 16 qualities (i.e., the sum of transmitted information for quality and intensity) did not yield improvement over quality alone. It was expected on the basis of the results of these experiments that the performance would be better than that for either intensity or quality alone but would fall short of the sum of the two. Analysis of individual judgment matrices indicated that very few errors were made on quality, but the intensity levels were badly confused. These confusions are most likely due to the fact that the three levels of intensity for each odorant were neither subjectively matched nor spaced on an equal discriminability scale. They were only equal in concentration dilution with the result that the strongest concentration dilution of Odorant A could be subjectively equal to the lowest concentration dilution of Odorant B, etc. In combining dimensions in such experiments one should deal with psychophysical rather than physical values. This seems especially important in the present case, but may also explain why, as information is added to sensory displays, the accuracy for any particular variable is so often drastically reduced (Miller, 1956). Just as there is an optimum spacing (jnds) of stimuli in a unidimensional experiment (Garner, 1953) one may expect there is an optimal combination of physical values in a multidimensional experiment.

SUMMARY

A series of experiments were carried out to determine how well the human *S* can identify odor qualities by label (S's own association to the odorant) with different kinds, intensities, and numbers of odorants. Informational analysis of these identifications yielded approximately 4.0 bits. The results were influenced by variations in the sample of odor qualities but not appreciably by intensity characteristics of the odor qualities. The information transmitted for a combination of 1.5 bits obtained for intensity earlier and the 4.0 bits obtained for quality did not exceed that for quality alone. The discussion of the results compared the span of absolute judgments in olfaction and other sense modalities with respect to unidimensional and multidimensional stimuli.

REFERENCES

ALLUISI, E. A. Conditions affecting the amount of information in absolute judgments. *Psychol. Rev.*, 1957, **64**, 97–103.

BECK, L. H., KRUGER, L., & CALABRESI, P. Observations on olfactory intensity: I. Training procedures, methods and data for two aliphatic homologous series. *Ann. NY Acad. Sci.*, 1954, **58**, 225–238.

BEEBE-CENTER, J. G., ROGERS, M. S., & O'CONNELL, D. N. Transmission of information about sucrose and saline solutions through the sense of taste. *J. Psychol.*, 1955, **39**, 157–160.

BEHNKE, A. R. Basic odor research correlation: Introduction. *Ann. NY Acad. Sci.*, 1954, **58**, 15–21.

ENGEN, T., & PFAFFMANN, C. Absolute judgments of odor intensity. *J. exp. Psychol.*, 1959, **58**, 23–26.

ERIKSEN, C. W. Multidimensional stimulus differences and accuracy of discrimination. *USAF, WADC Tech. Rep.*, 1954, No. 54–165.

GARNER, W. R., & HAKE, H. S. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.

KLEMMER, E. T., & FRICK, F. C. Assimilation of information from dot and matrix patterns. *J. exp. Psychol.*, 1953, **45**, 15–19.

McGILL, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97–116.

MILLER, G. A. The magical number seven, plus or minus two: Some limitations on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

POLLACK, I. The information of elementary auditory displays. II. *J. Acoust. Soc. Amer.*, 1953, **25**, 765–769.

POLLACK, I., & FICKS, L. Information of elementary multi-dimensional auditory displays. *J. Acoust. Soc. Amer.*, 1954, **26**, 155–158.

(Early publication received December 29, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

ORDER OF CONSIDERATION OF DIFFERENT TYPES OF CONCEPTS[1]

EARL B. HUNT[2] AND CARL I. HOVLAND

Yale University

In the learning of some types of concepts individuals acquire the correct classification without any conscious or clear-cut plan of attack. This is typical of much of our perceptual learning and even of many simple conceptual tasks. But in other types of situations, for example those that are involved in scientific analysis, individuals characteristically go about the task of finding the concept by some type of deliberate strategy. This was the most common type of approach in the acquisition of concepts involving geometrical figures in the studies of Hovland and Weiss (1953) and of Bruner, Goodnow, and Austin (1956). These studies fostered deliberate approaches because of the tasks chosen and the nature of the instructions. For such problems one can logically distinguish a number of different kinds of concepts, derived from the types of classifications of instances described by the logical operations "and" and "or." In the *conjunctive* type all of the instances have features in common, so the concept is one where each instance possesses Characteristics A *and* B, or A *and* B *and* C, etc. In the case of *disjunctive* concepts, all instances have one *or* another feature (cf. Bruner, Goodnow, and Austin's [1956] example of the concept "strike" in baseball, where a strike is defined as either a pitched ball which crosses the plate between the batter's knees and his shoulders *or* is alternatively any pitch struck at by the batter which fails to be sent into the field). The authors just mentioned found this type of concept exceedingly difficult to learn. The third type of concept, discussed most extensively by Smoke (1935), is *relational*, in which the common properties are sets of relationships rather than common specific stimulus elements. For example, the concept of isosceles triangle involves the common relationship of equality between two of the sides, and the positive instances may have widely differing size and shape.

Subjects typically structure a concept problem in one or another of these ways and sometimes pursue their initial formulation until the evidence for its being unsuitable for the problem becomes overwhelming. In machine simulation of human concept learning it then becomes important to know the order in which solutions of each type will be tried (cf. Hovland, 1959). Since there may be individual differences in the order in which alternatives are considered it may also become necessary to establish the frequency of appearance of a particular order for different segments of the total population.

Order of attack may also be influenced by the perceptual characteristics of the stimuli to be classified. Within a single type of concept (conjunctive), variations in order of consideration have been demonstrated by Heidbreder and her associates (1945) and by Grant (1951). The *S*s in these experiments first tried to classify on the basis of form; when this was unsuccessful they tried color and, finally, number of figures. One might also anticipate an interaction between stimulus characteristics and type of concept selected such that when certain dimensions are used conjunctive concepts would be considered first whereas with others disjunctive or relational concepts would be tried first.

---

[1] This research was financed by a Ford Foundation grant to Carl I. Hovland for the investigation of concept formation: this support is gratefully acknowledged.

[2] The present experiment was performed while this author held a General Electric Educational and Charitable Fund Fellowship at Yale University.

The present experiment was designed to obtain a systematic comparison of the frequency with which *S*s choose each of the three concept types as their initial basis for organizing instances. A series of geometrical forms labeled as positive or negative instances was presented to *S*. The instances were chosen in such a way that they could be organized with equal logical validity on a conjunctive, disjunctive, or relational basis. Which type of concept *S* utilized was determined on the basis of which of a series of test instances he selected as positive instances of the concept which he had derived from the training series. Each *S* was given one of three different series of positive and negative instances in which color, form, and number dimensions of the stimuli were rotated as appropriate bases for particular types of concept. This design made it possible to evaluate any interaction between stimulus characteristics and type of concept utilized.

METHOD

Stimuli. The stimuli were designs of the type illustrated in Fig. 1. The figures were in purple ditto ink; the stripes in the upper left and lower right corners were colored by crayon.

The stimuli involved six dimensions: color of the top stripe, color of the bottom stripe, number and form of the upper figure, number and form of the lower figure. The values used were: color (red, green, light brown, black); number (one through four); figure (triangle, cross, fleur-de-lis, and star). The sets of 15 stimuli each used during training were reproduced on 8½ x 11-in. sheets; two contained six instances each, the third three. Assignment of stimuli to a particular location and page was random. The 16 similar stimuli used in testing were presented eight to a page.

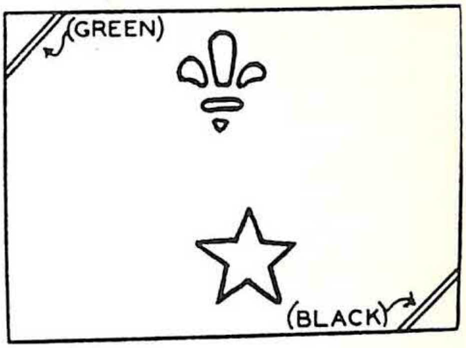


FIG. 1. Illustrative stimulus pattern.

Design. Each *S* was randomly assigned to one of three conditions corresponding to the set of stimuli to be used in training. During training each *S* received a set of positive or negative instances from which one conjunctive, disjunctive, or relational concept could be derived. The three training sets provided rotation of the dimensions used for each type of concept. Following the training set *S* selected concept exemplars from a test set consisting of unlabeled stimuli.

Each training set contained nine positive instances (labeled "Alpha") and six negative instances (labeled "not Alpha"). These completely defined one conjunctive, one relational, and one disjunctive concept, provided *S*'s hypotheses involved less than three dimensions.[3] The sets of instances used in each of the three training series are presented in Table 1. It will be seen that in Training Series I the conjunctive concept "upper red stripe and lower black stripe" would be consistent with this set of positive and negative instances. The relational concept "same number of upper and lower

---

[3] Examination of the pattern of selection of test instances by *S*s showed that *S*s did not use categorizing rules based on three or more dimensions.

TABLE 1

DESCRIPTION OF STIMULI[a]

	Set I						Set II						Set III					
	Upper Half			Lower Half			Upper Half			Lower Half			Upper Half			Lower Half		
	Fig.	No.	Stripe Color	Fig.	No.	Stripe Color	Fig.	No.	Stripe Color	Fig.	No.	Stripe Color	Fig.	No.	Stripe Color	Fig.	No.	Stripe Color
Training: Positive Instances																		
1.	C	4	R	F	4	Bk	C	2	G	C	3	Br	S	1	Bk	T	4	Bk
2.	T	4	R	F	4	Bk	C	2	Bk	C	3	Br	S	3	Bk	T	4	Bk
3.	C	1	R	C	1	Bk	S	2	G	S	3	R	S	1	R	T	1	R
4.	C	1	R	T	1	Bk	S	2	G	S	3	Br	S	1	R	T	4	R
5.	C	3	R	S	3	Bk	F	2	G	F	3	G	S	1	Br	T	3	Br
6.	C	2	R	S	2	Bk	T	2	G	T	3	Br	S	1	G	T	4	G
7.	C	2	R	T	2	Bk	T	2	G	T	3	Bk	S	1	G	T	3	G
8.	F	3	R	F	3	Bk	F	2	R	F	3	Br	S	2	Br	T	4	Br
9.	S	1	R	F	1	Bk	S	2	Br	S	3	Br	S	4	R	T	4	R
Training: Negative Instances																		
1.	T	1	R	S	4	G	S	2	Bk	C	4	Gr	S	2	R	S	1	Bk
2.	T	2	G	T	4	Bk	T	4	Bk	C	3	Bk	C	2	Gr	T	1	Bk
3.	S	2	R	T	3	Br	T	2	Br	F	2	Bk	S	3	Gr	F	2	Br
4.	S	1	Br	S	3	Bk	S	3	Br	F	3	G	F	3	R	T	2	Br
5.	F	3	R	C	2	G	F	2	R	S	1	R	S	4	Br	C	3	G
6.	F	4	Bk	C	1	Bk	C	1	R	T	3	R	T	4	Bk	T	3	R
Test Instances																		
A.	F	1	G	S	1	Bk	(Test instances same for all groups)											
B.	T	2	Br	S	3	G												
C.	S	2	Bk	T	2	Br												
D.	S	1	R	S	4	Bk												
E.	S	3	R	T	2	Bk												
F.	C	4	R	C	2	R												
G.	T	4	R	C	4	Br												
H.	S	3	Br	T	3	G												
I.	T	1	R	T	4	Bk												
J.	S	4	R	T	1	Bk												
K.	F	2	Bk	S	3	Bk												
L.	T	2	R	C	3	R												
M.	C	4	G	F	3	Br												
N.	F	3	Bk	F	1	Bk												
O.	C	2	G	F	1	Br												
P.	F	2	Bk	C	3	R												

[a] In the training set, positive and negative instances were interspersed randomly. The description code for the figures is S—star, C—cross, F—fleur-de-lis, T—triangle. For colors it is R—Red, G—Green, Br—Brown, Bk—Black.

figures" would be equally correct, as would the disjunctive concept "crosses in the top row and/or fleur-de-lis in the bottom row."

Following presentation of the training series *S*s were given the test set: 16 stimuli of the same type, but without labels. A description of these stimuli is given in the lower portion of Table 1. This set was the same for all *S*s. Subjects were instructed to pick out those instances in the test set which they considered to be positive instances of the concept they had derived from the training set. One set of four instances would be selected if *S* had derived a conjunctive concept from the training series, another set of four if he had utilized a disjunctive one and a third set of four if it were relational. If the reader directs his attention to the relationship between the test set and Training Series I he will see that four of the patterns (D, E, I, J) are further instances of the conjunctive concept of this training series, four others (A, C, G, H) are positive instances of the relational concept, and four others (F, M, N, O) are positive instances by the disjunctive definition. In addition four "dummy" instances (B, K, L, P) were included which could not be considered as positive instances by any of the three categorizing rules derivable from the training series. Similar relationships between training and test stimuli obtained for the other two series.

The *S*s could choose as many "positive" instances as they wished. If the majority of the test stimuli selected by *S* were consistent with the conjunctive definition he was categorized as having utilized a conjunctive concept; a similar categorization was applied for relational and disjunctive concepts. If the majority of *S*'s choices were of "dummy" instances or if he did not place a majority of his choices in one of the other three categories he was classified as "inconsistent."

Subjects. Thirty-seven students from the New Haven College day school and 25 from the night school served as paid *S*s.[4] One *S* in each group had to be dropped for failure to understand the instructions. No *S* had previously participated in a concept formation experiment.

Procedure. Each *S* was handed an instruction sheet and an envelope containing the test and training series. The *E* read the instructions, answered specific questions, and described the six dimensions. The *S* then read the training sheets, returned them to the envelope, took out the test sheets, selected his choice of instances, and handed in the material. This procedure made it possible to present all training sets to the group and to randomly assign *S*s to each training set.

RESULTS

The *S*'s utilization of a particular concept type was evaluated in terms of the number of stimuli of each type which he indicated as exemplars. As stated above, each *S* could select from zero to four test instances which would be consistent with a particular concept derivable from the training set. The experiment can be analyzed as a 2 × 3 × 4 "mixed" design corresponding to Lindquist's (1953) Type III. Each *S* is involved in one "treatment" of schools. one of "training set" and all measures for the learning of a particular concept type. In this analysis it is necessary to have proportional numbers of *S*s in each of the three conditions (types of training sets) across schools. For this reason the number of cases for each training series within schools was randomly reduced to 11 (for the day school) and seven (for the night school). Analysis of variance methods were used although the scores within each treatment were not normally distributed and the range of scores was curtailed. This makes a conservative interpretation of significance levels mandatory. It will be seen that the differences presented in Table 2 are large. There is no evidence that the two school populations differed on the measures taken.

TABLE 2

ANALYSIS OF VARIANCE OF CHOICES OF INSTANCES ON THE TEST SERIES

Source	*df*	*MS*	*F*
Between *S*s	53		
Schools	1	.26	.20
Training sets	2	7.59	5.97*
Stimulus × Schools	2	.83	.65
Error (between)	48	1.27	
Within *S*s	162		
Concepts	3	32.01	14.28*
(Relational and conjunctive) vs. (Disjunctive and error)	1	96.00	42.84*
Relational vs. conjunctive	1	.02	.01
Disjunctive vs. error	1	.02	.01
Concepts × Schools	3	.70	.31
Concepts × Training Sets	6	.74	.33
Concepts × Schools × Training Sets	6	3.82	1.70
Error (within)	144	2.24	

* $P < .01$.

---

[4] The authors wish to express their thanks to the administrative and teaching staff of New Haven College for their cooperation, and to Benton S. Underwood and Fred Sheffield for helpful suggestions for the writing of this report.

TABLE 3

MEAN NUMBER OF INSTANCES CHOSEN FROM THE TEST SERIES

Training Set	Type				Mean Total
	Relational	Conjunctive	Disjunctive	"Inconsistent"	
	Day School ($N = 33$)				
I	1.73	2.00	1.09	.91	5.73
II	1.46	1.82	.09	.00	3.37
III	1.82	1.27	.45	.36	3.90
	Night School ($N = 21$)				
I	3.43	1.71	.43	.29	5.86
II	1.43	1.29	.00	.14	2.86
III	1.14	2.56	.57	.86	5.14

Mean total by training set: I = 5.78, II = 3.17, III = 4.39.
Mean total by schools: Day = 4.33, Night = 4.57.

The mean numbers of instances selected by *S*s, by concept type and training set are presented in Table 3. For the three training sets there was a tendency to check different numbers of instances. Possibly this related to problem difficulty and the degree of certainty felt by *S*s after they had seen a particular training set. It will be remembered that an *S* who learns, with absolute certainty, one categorizing rule will check exactly four instances. The major part of the variance was due to *S*s concentrating their choices on relational and conjunctive concept exemplars much more frequently than any other type of instance in the test series.

The frequency of choice of particular types of concept exemplars did not vary with training set, as is shown by the nonsignificant Concepts × Training Set interaction.

The analysis of mean number of instances of each type of concept chosen by *S*s is supported by an examination of *S*s' concentration of choices of instances which are consistent with a particular type of concept derivable from the training set. Four *S*s concentrated their choices on disjunctive instances, eight on "dummy" instances (or had no majority), 24 concentrated on conjunctive choices and 24 concentrated on relational choices. There was no difference in frequency of choice concentrations between training sets or between the day- and night-school *S*s.

DISCUSSION

The above results provide a tentative answer to the question concerning the order in which different types of concepts are considered. Under our experimental conditions both conjunctive and relational concepts are selected significantly more frequently than disjunctive ones. Accordingly, we plan to program our initial computer simulations so that conjunctive possibilities are considered alternately with relational ones. Only when both have been explored without success will the program of the computer consider possible solutions involving disjunctive concepts. The present experiment does not provide information as to what determines an *S*'s initial choice of conjunctive or relational concept. This information will be needed to specify which of two alternative programs should be applied to simulate the behavior of a particular individual on a given concept learning problem.

A number of limitations on the generality of the present results must be considered. One concerns the population of concepts from which the limited set used in the present experiment was derived. Are our relational concepts, for example, typical of the population of such concepts or is there some tendency for investigators to select ones which are particularly easy to learn? And are our results in which "sameness" was used for the relational concept equally applicable to relationships such as "subordinate to" or "bigger than"?

There may be some question as to whether the three types of concepts are distinctive psychologically. For example, some *S*s use *and* and *or* interchangeably, some describing a particular class of stimuli as consisting of red *or* green triangles while others describe the same grouping as red *and* green triangles. Similarly, when pairs of large and small circles and triangles are the stimuli some derive the concept of "two large triangles *or* two small triangles" (a disjunctive concept) while others speak of a grouping in which "the triangles are both the same size" (a relational concept).

Another restriction on generality may arise from the use of college students in the present experiment. It is possible that the preference for conjunctive and relational concepts is a product of learning and that children may tend to select disjunctive concepts more frequently than adults.

No evidence was obtained of interaction between the perceptual characteristics of the stimuli and the type of concept chosen, although interaction effects had been anticipated. It is felt that further exploration of this problem is indicated with other quite different types of stimulus material, including some where the dimensions are more difficult to discriminate. But until evidence for interaction is obtained it does not appear necessary to provide alternative computer programs for different types of stimuli in simulation studies.

SUMMARY

The *S*s in a visual concept formation task could derive a correct concept from the instances presented on a conjunctive, disjunctive, or relational basis. It was found that conjunctive and relational solutions were utilized much more frequently than disjunctive ones. No significant interaction effects were obtained between the particular aspects of the stimulus involved (color, number, or type of figure) and the frequency of utilization of the different types of concept. The experiment is discussed in relation to machine simulation of human concept learning.

REFERENCES

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

GRANT, D. A. Perceptual versus analytic responses to the number concept of a Weigl-type card sorting test. *J. exp. Psychol.*, 1951, **41**, 23–29.

HEIDBREDER, E. Towards a dynamic psychology of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.

HOVLAND, C. I. Computer simulation techniques in behavioral science research. Talk given on *Voice of America*, 1959. (Mimeo.)

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

LINDQUIST, E. F. *The design of experiments in psychology and education.* Boston: Houghton-Mifflin, 1953.

SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**, (4, Whole No. 191).

(Early publication received
October 30, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

REINFORCERS AND REINFORCEMENT: THEIR RELATION TO MAZE PERFORMANCE[1]

WILLIAM BEVAN AND ROBERT ADAMSON[2]

Emory University

Thinking about reinforcement has been marked, for the most part, by a tendency to equate internal process and stimulus agent and a preoccupation with frequency of presentation of the agent as the crucial reinforcement variable. Thus Tolman and Postman (1954), have commented, ". . . it is our view that it is the frequency of S-R-S occurrences and not amounts of reinforcement which primarily affect acquisition" (p. 37). Yet the appearance of more than 25 papers on reinforcer magnitude in the past several years challenges this opinion. A cursory examination of this literature clarifies the need for distinguishing between process and agent. While older studies, like those of Gantt (1936) and Grindley (1929), have indicated a positive connection between performance efficiency and incentive magnitude, the demonstration of contrast effects (e.g. Crespi, 1942; Zeaman, 1949) denies the possibility that the reinforcer-performance relationship is simple and constant. Estes (1956), in referring to the difficulty in generalizing about the nature of this relationship, has concluded that progress toward systematization depends upon the construction of a common scale of measurement. A fruitful approach to this requirement he sees in a ". . . correlation of measures obtained in learning experiments with those obtained in psychophysical experiments" (p. 13). Previous to this, Campbell and Kraeling (1953), studying noxious stimulus reduction as reinforcement, had confirmed the expectation that reduction of shock by a given amount had less effect upon performance with high levels than with low levels of stimulation. They had proposed that some function like Weber's law may be applicable in describing reinforcement.

The concern of the present writers with the magnitude problem derives from an interest in the range of applicability of the approach of frame-of-reference psychophysics. This orientation has provided for the competent prediction of sensory magnitude from an assessment of the intensities of a sequence of stimulus inputs. It has also effectively accounted for anchor and context effects, constancy and contrast, and certain aspects of stimulus equivalence (Helson, 1959). Should it be permissible to assume a reinforcement analogue for the sensory adaptation level, shifts in the effectiveness of reinforcers, contrast effects, the retroaction paradox, adaptation to repeated reinforcement, the superiority of partial reinforcement, and other phenomena associated with performance efficiency might be brought under a single rubric. Meanwhile, examination of the appropriateness of applying the adaptation-level model in the reinforcement context requires first the demonstration that reinforcement has certain properties. The experiments of the present paper were directed toward identifying several of these prerequisite characteristics.

---

[1] The experiments of the present paper were conducted during the summer of 1957 and the fall of 1958 under a grant from the Emory University Research Committee. The writers would like to thank Harry Barker and James M. Burt for their assistance in data collection. Systematization of the data was part of a project carried out during the summer of 1958, while the writers were in attendance at the Interdisciplinary Behavioral Sciences Research Conference, University of New Mexico. This conference was supported by the Behavioral Sciences Division, Air Force Office of Scientific Research, under Contract AF 49 (638)-33.

[2] William Bevan is now at Kansas State University, Manhattan, Kansas; Robert Adamson is at System Development Corporation, Lodi, New Jersey.

EXPERIMENT I: SCALING OF ELECTRIC SHOCK

Adaptation-level theory describes apparent sensory magnitude as a differential between a present stimulus process and an internal norm derived, through a process of pooling, from background and residual (past) stimulation. In this initial experiment, the primary goal was to determine if electric shock, a widely used reinforcer, was scalable in the same fashion as prothetic dimensions to which AL theory already has successfully applied. That shock is scalable is clear. Prior to the present experiment, Swartz (1953) had used shock to the tooth pulp to construct a fractionation scale for pain; and more recently, Stevens, Carton, and Shickman (1958) have described both category and ratio scales for electric shock. The present study, however, is concerned less with delineating the scalar properties of shock, than with demonstrating the generation of an indifference point (AL) dependent upon the relative frequency of the several magnitudes of the stimulus series. This is a requirement of Ex. III.

Procedure. Thirty-seven undergraduate and graduate students each made 100 judgments daily on three consecutive days. For each series of 100 shocks, there were five discrete intensities delivered by a square-wave generator (Applegate stimulator, Model 226B): 1300, 1800, 2300, 2800, and 3300 μamp ac, with a pulse rate of 100 cps. Duration of each presentation, controlled by a Wichita stimulus controller, was 200 msec. The interval between successive presentations was 10 sec. The onset of each shock, manually controlled, was signaled by a small green light, turned on by a Phipps and Bird synchronous clock. Shock electrodes were two 2-cm. silver discs coated with bentonite paste and attached to the left wrist, one on the proximal and one on the distal side.

TABLE 1

AVERAGES FOR EACH STIMULUS SERIES, BASED ON THREE ASSUMED SHAPES OF PSYCHOPHYSICAL FUNCTION, AND THE CORRESPONDING EMPIRICALLY OBTAINED INDIFFERENCE POINT

Order	Arithmetic Mean	Geometric Mean	Parabolic Mean	Indifference Point
A	2300	2184	2395	2287
B	2050	1947	2109	2087
C	1900	1817	1947	1900

The psychophysical method was the rating scale version of the absolute method, with nine categories from *very, very weak* through *medium* to *very, very strong*. The usual procedural precautions were observed, *S* being requested to base his judgments only on the shocks presented in the current session. The *S*s were allowed additional categories at either end of the rating scale.

For each of the three sessions, a different order of presentation was used: a rectilinear-control series (A), consisting of each of the five intensities presented 20 times in random order; a positively skewed series (B) with the intensities—weakest to strongest—presented 30, 25, 20, 15, and 10 times, respectively; and another positively skewed series (C) with intensities presented 35, 30, 20, 10, and 5 times, respectively. These orders were presented in the following sequences: ABC, BCA, and CAB, approximately one third of the *S*s being randomly assigned to each sequence. The Indifference Point (median category judgment) assigned to each intensity for each *S* under each condition was computed and averaged through *S*s. The data for each series were fitted by rectilinear, logarithmic, and parabolic functions and the appropriate average for each condition determined. These are presented in Table 1, the right-most column containing the empirically obtained IPs for the several series.

Results. Inspection of Table 1 indicates, first, that the judgmental responses to shock, like the prothetic dimensions previously subjected to category judgment, are scalable and yielded IPs. Secondly, the IPs for the positively skewed series were less than for the rectilinear series and in the expected order (between-series $F = 9.69$; $df = 2/108$; $P < .001$), indicating the induction of anchor effects. Thirdly, the IPs correspond most closely to the arithmetic means of the stimulus series suggesting the S-R function to be linear. This agrees with Swartz's (1953) report of an approximately linear pain function for stimulation to the tooth pulp. Stevens, Carton, and Shickman (1958) also report that their category data for shock to the fingers could be fit by a straight line but argue, because of the nonlinear relationship between their category and magnitude-estimation data, that this is not indicative of direct proportionality between subjective intensity and shock.

EXPERIMENT II: SUBJECTIVE INTENSITY AND MAZE PERFORMANCE

Extension of the adaptation-level model to reinforcement requires the demonstration that performance efficiency is a function of the apparent rather than the physical intensity of the reinforcing agent. The contrast effects reported in a number of animal experiments with positive reinforcers strongly suggest this to be the case. Both Crespi (1942) and Zeaman (1949) indicate running speed for small amounts of food to drop below an expected level when their *S*s had been previously run to a larger amount, while the opposite held for animals initially run to a small amount and shifted to a larger reward. More recently, Collier and Marx (1959) have reported similar differences in acquisition curves for bar pressing. But the contrast data are ambiguous with regard to the role of subjective intensity, since the criterion trials are, in most studies, run with different physical magnitudes of reinforcer. Further, all have involved changing level of reinforcement during the sequence of acquisition trials, allowing for the possibility that shift may be the significant variable, and thus making plausible explanation in more directly motivational terms. Indeed, Crespi, for example, has described his results in terms of the rodent analogues of elation and depression. Meanwhile, Collier and Marx tested with a common intermediate magnitude. They also adapted with magazine training and tested with bar pressing. The present experiment involves comparing maze learning to a common intensity of shock, after pretest adaptation with different intensities of shock, applied outside the context of the learning situation.

Procedure. Three groups of 17 *S*s were required to judge the intensity of 30 shocks in a pretest situation, using the rating scale of Exp. I. Group A received 30 shocks of 1300 μamp at 10-sec. intervals, Group B, the control group, 30 of 2300 μamp., and Group C, 30 of 3300 μamp. Following the judging series, all *S*s were given a 28-unit, two-choice, bolthead maze problem and required to make five trips around the maze perimeter. Each trip or trial sequence thus involved 28 choices. The maze stylus was manipulated by the right hand, contact with the incorrect bolthead of each pair resulting in the delivery of a single 200-msec. shock of 2300 μamp. through electrodes on the left wrist. The instrumentation for the automatic delivery of reinforcement, as well as for the required programming of shocks of different intensities is described elsewhere (Adamson & Bevan, in press). Choices were spaced at 10-sec. intervals, *S* being paced by the green signal light used in Exp. I. The dependent variable consisted of mean errors for each of the successive trial sequences. These are graphed in Fig. 1 and were subjected to a series of trend analyses (Alexander, 1946) from which the comparisons of interest are summarized in Table 2A.

Results. Inspection of Fig. 1 indicates that maze performance differs with the apparent intensity of reinforcing shock. On the initial sequence, all three groups performed near a chance level, but with subsequent trips the curves diverge, indicating different rates of acquisition. Acquisition was most rapid in Group A, for which the reinforcement was strong, relative to the magnitude of adapting shock. In contrast, performance was poorest in Group C, for which reinforcement was relatively weak. The over-all trend test performed on all three groups (cf. Table 2A) yielded significant *F*s for both between-group means and between-group slopes sources of variance. Tests performed on the several-paired combinations of the groups indicate Group C to be inferior to both B and A, but B not to differ from A. It is possible that had the difference between adapting and test shock levels been greater for A, its superiority over B might have been statistically as well as graphically demonstrated.

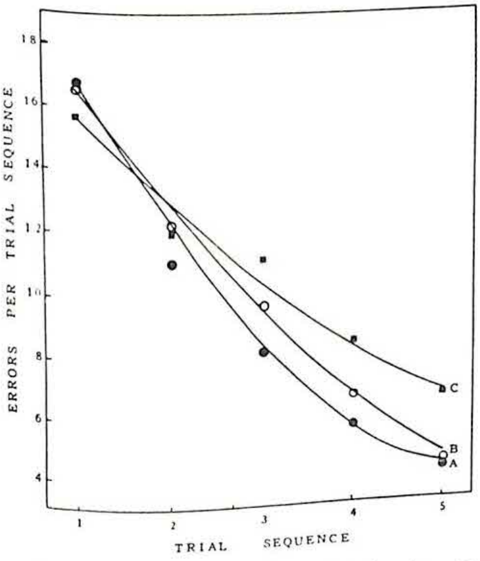


FIG. 1. Performance on a bolthead maze as a function of the relative intensity of reinforcing shock to pretest adaptation shock. For Group A the reinforcing shock was more intense than the pretest shock. For Group B, both were of the same intensity. For Group C the reinforcing shock was weaker. For all groups the reinforcing shock was of the same physical intensity.

TABLE 2

F RATIOS OBTAINED FROM TREND TESTS: EXP. II AND III

Comparison	*df*	*F*	*P*
A. Experiment II			
Over-all Means	2/144	4.96	<.01
Over-all Slopes	2/144	5.81	<.01
C–B Means	1/96	4.05	<.05
C–B Slopes	1/96	7.95	<.01
C–A Means	1/96	15.61	<.01
C–A Slopes	1/96	9.25	<.01
B–A Means	1/96	2.57	>.05
B–A Slopes	1/96	0.001	>.05
B. Experiment III			
Over-all Means	2/135	44.00	<.001
Over-all Slopes	2/135	10.38	<.001
A–B Means	1/100	29.00	<.001
A–B Slopes	1/100	165.00	<.001
A–(C, C′) Means	1/108	86.80	<.001
A–(C, C′) Slopes	1/108	11.80	<.001
B–(C, C′) Means	1/108	8.96	<.001
B–(C, C′) Slopes	1/108	3.66	<.05
C–C′ Means	1/78	.093	>.05
C–C′ Slopes	1/78	.002	>.05

EXPERIMENT III: AVERAGING OF REINFORCEMENT OVER TRIALS

Adaptation-level theory assumes that the internal referent that determines subjective magnitude is the product of pooled residual and background stimulation. After determining the relative influence of residual and background for particular sensory dimensions and psychophysical methods, it is possible to quantitatively specify shifts in the Indifference Point with changes in the relative frequency of presentation of the several series stimuli and with the introduction of anchors (Helson, 1947; Johnson, 1944). It is also possible to predict equivalent patterns of input (Bevan & Darby, 1955). The analogue for the reinforcement concept would involve predicting the magnitude of change in performance efficiency from changes in the magnitudes comprising a series of reinforcing stimuli as well as identifying a priori equivalent magnitude patterns of reinforcer intensities.

Prerequisite to this is the demonstration that performance efficiency may be related systematically to average reinforcement level. Supportive evidence is found in the report of Logan, Beier, and Ellis (1955) that average running speed did not differ between a group reinforced with five pellets on each trial and a group receiving nine pellets on one half of the trials and one on the alternative half. Also relevant are the findings of Leventhal, Morrell, Morgan, and Perkins (1959) that no preference could be demonstrated for an alley reinforced on every trial over one reinforced with double the amount on half the trials. In the present experiment, the maze performance of a control group receiving a fixed magnitude of shock for every error is compared to that of a group receiving a symmetrical distribution of shocks with the same mean intensity and to groups receiving positively and negatively skewed distributions of shock. It is expected that the first experimental group will not differ significantly from the control, while those receiving greater and less average shock will perform with different efficiency.

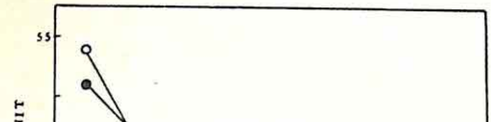


FIG. 2. Performance on a bolthead maze as a function of the mean intensity of reinforcing shock. Group A received the five intensities used in the first experiment, skewed from weak to strong, in the following proportions: 31%, 26%, 22%, 14%, and 7%. Group B received the same intensities, skewed from weak to strong, in these proportions: 7%, 14%, 22%, 26%, 31%. Group C received a symmetrical distribution about the mean intensity in the following proportions: 7%, 22%, 42%, 22%, and 7%. Group C′ received a single intensity (the mean value) on *all* reinforced trials.

Procedure. In order to separate the possible effects of varying intensity over trials from varying frequency of reinforcement, it was necessary to insure that all *S*s receive the same number of reinforced trials. Accordingly, a more difficult task—a 40-unit, two-choice, bolthead maze problem—was used to insure a specified minimum of reinforced trials. Four groups of *S*s, 48 in all, were tested. The five shock intensities used in Exp. I were employed in various combinations as reinforcers. Group A received, on incorrect trials, a positively skewed random distribution of 70 shocks as follows: 31%, 26%, 22%, 14%, and 7%, weakest to strongest, the mean intensity being 2000 μamp. Group B received a negatively skewed distribution, mirroring that of Group A, with a mean intensity of 2600 μamp. Group C received a symmetrical distribution (7%, 22%, 42%, 22%, and 7%) about a mean of 2300 μamp., and Group C′ received a single intensity of 2300 μamp. for all errors. The test procedure and apparatus were the same as that for Exp. II except that no pretest shocks were given. Performance was scored as follows: The total number of choices made to the 70-error criterion was counted, then divided into fifths, each fifth being designated as an error unit. The relative frequency of errors per error unit was then computed, and the data subjected to the Alexander test for trend. The results are presented in Fig. 2 and Table 2B.

Results. Examination of Fig. 2 suggests that Group A (lowest mean intensity) gave the poorest performance and Group B (highest mean intensity) the next poorest. The best performance was associated with the average shock of intermediate intensity. Meanwhile, Groups C and C′, in line with expectation, did not differ significantly from each other. They were thus combined for statistical comparison with Groups A and B. Table 2B indicates each of the three mean reinforcement magnitudes to differ from the remaining two. It would appear, therefore, that performance level relates reliably to average reinforcement level. The establishment of this relationship is necessary to the extension of the pooling model to this problem area. That an averaging effect is tantamount to the pooling of reinforcement over time is, however, still unclear. A definitive demonstration of the latter requires the quantitative prediction of performance associated with some designated average level from the performance levels associated with each of the component intensities that contribute to the average. The present writers currently are conducting a parametric study of runway behavior, with reinforcer magnitudes of 1, 2, 3, or 4 pellets on 25, 50, 75, or 100% of the trials, as a preliminary to attempting such a prediction. Meanwhile, the operation of a pooling mechanism in the present situation would be suggested by a lack of correlation between repetitive errors at specific choice points and specific strengths of shock. Such data are unfortunately not presently available. However, an ancillary study, which indicated a low but significant negative correlation between the number of perseverative errors on maze performance and the magnitude of AL shifts in the judgmental situation, failed to reveal any consistently missed choice points over individuals. Furthermore, four *S*s of Group C′ achieved errorless performance at the end of the second trial sequence, and thus had only a minimum opportunity for the perseveration of errors. Finally, a sizable proportion of *S*s stated that they remembered their strongly shocked responses best, while an equally large number reported that the more intense shocks confused them.

Discussion

The extension of a psychophysical model to the area of learning and performance has several distinct values. For example, the pooling model—through assessing the value of the reinforcing agent in terms of its discrepancy from an average—can readily account for the phenomenon of contrast. Similarly, the adaptation effect, associated with repeated presentations of a single-intensity reinforcer, may be explained as follows: If it may be assumed that, prior to the initial presentation of a reinforcer, the magnitude of the internal norm is zero, then, after the first presentation, the norm will assume a value intermediate between zero and the value of the reinforcing agent. With successive presentations of a fixed intensity, the value of the norm will more and more closely approximate the reinforcer, and the discrepancy between the norm and the reinforcing agent will diminish, as will the apparent intensity of the reinforcer. Conversely, extinction trials should have the effect of reducing the norms. If then reinforcing stimuli are reintroduced, their value should be enhanced. Reports that postextinction performance exceeds a pre-extinction asymptote (Adamson, Bevan, & Maier, 1959; Lauer & Carterette, 1957) support this interpretation. The approach applied to the adaptation effect is also readily applicable to the problem of partial reinforcement. Furthermore, the retroaction paradox disappears when reinforcement is thought of as temporally extended.

Summary

The present paper describes three experiments directed toward examining the status

of three assumptions prerequisite to the extension of the pooling model, associated with adaption-level theory, to the area of reinforcement. The assumptions were: (*a*) that *S*s, presented a common reinforcing agent as a psychophysical stimulus, would behave as they do when faced with the task of scaling a prothetic dimension—intensities would be ordered, PSEs derived, and anchor effects in evidence; (*b*) that maze performance could be related in significant degree to the relative, in contrast to the absolute, intensity of the reinforcer; and (*c*) that maze performance would vary as a function of the average intensity of reinforcer.

The *S*s were 136 college students given electric shocks of weak-to-mild intensity either as stimuli for absolute judgments or following errors on a bolthead maze. The results of all three experiments confirmed the validity of the assumptions.

REFERENCES

ADAMSON, R. E., & BEVAN, W. A simple and inexpensive programming device for psychophysical and learning experiments. *J. gen. Psychol.*, in press.

ADAMSON, R. E., BEVAN, W., & MAIER, B. Shifts in bar pressing as a function of alternating extinction and reinforcement. *J. gen. Psychol.*, in press.

ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.

BEVAN, W., & DARBY, C. L. Patterns of experience and the constancy of an indifference point for perceived weight. *Amer. J. Psychol.*, 1955, **68**, 575–584.

CAMPBELL, B. A., & KRAELING, D. Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, 1953, **45**, 97–101.

COLLIER, G., & MARX, M. H. Changes in performance as a function of shifts in the magnitude of reinforcement. *J. exp. Psychol.*, 1959, **57**, 305–309.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

ESTES, W. K. Learning. *Ann. Rev. Psychol.*, 1956, **7**, 1–38.

GANTT, W. H. An experimental approach to psychiatry. *Amer. J. Psychiat.*, 1936, **92**, 1007–1021.

GRINDLEY, G. C. Experiments on the influence of amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, **20**, 173–180.

HELSON, H. Adaption-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HELSON, H. Adaptation-level theory. In S. Koch, (Ed.), *Psychology: A study of a science.* Vol. I. New York: McGraw-Hill, 1959. Pp. 565–621.

JOHNSON, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, **34**, 425–436.

LEVENTHAL, A. M., MORRELL, R. F., MORGAN, E. F., JR., & PERKINS, C. C., JR. The relation between mean reward and mean reinforcement. *J. exp. Psychol.*, 1959, **57**, 284–287.

LOGAN, F. A., BEIER, E. M., & ELLIS, R. A. Effect of varied reinforcement on speed of locomotion. *J. exp. Psychol.*, 1955, **49**, 260–267.

LAUER, D. W., & CARTERETTE, T. S. Changes in response measures over repeated acquisitions and extinctions of a running habit. *J. comp. physiol. Psychol.*, 1957, **50**, 334–338.

STEVENS, S. S., CARTON, A. S., & SHICKMAN, G. M. A scale of apparent intensity of electric shock. *J. exp. Psychol.*, 1958, **56**, 328–334.

SWARTZ, P. A new method for scaling pain. *J. exp. Psychol.*, 1953, **45**, 288–293.

TOLMAN, E. C., & POSTMAN, L. Learning. *Ann. Rev. Psychol.*, 1954, **5**, 27–56.

ZEAMAN, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received April 30, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

RELATIONSHIP BETWEEN MAS SCORES AND ASSOCIATION VALUES OF NONSENSE SYLLABLES[1]

E. PHILIP TRAPP AND DONALD H. KAUSLER

University of Arkansas

In studies relating anxiety drive level, as measured by the Taylor Manifest Anxiety Scale (MAS), to performance on a verbal learning task, a frequently used material has been nonsense syllables, presented either serially (e.g., Montague, 1953; Saltz & Hoehn, 1957; Sarason, 1956, 1957; Silverman & Blitz, 1956) or as paired associates (e.g., Taylor & Chapman, 1955). A procedure commonly followed in these studies is to select syllables systematically in order to control for task difficulty and/or intratask competition. The specific variables generally manipulated for this purpose are the similarity of the syllables within the list and the association value of these syllables. Similarity is controlled by means of the letter content of the syllables, and association value is usually determined from one of the early calibrations by Glaze (1928) or Hull (1933). The selected syllables then become the learning task for the groups differentiated on the anxiety dimension.

The present study is concerned with association value since it, like MAS score, is an *S* variable. Similarity is of perhaps equal importance as a variable in determining task difficulty and intratask competition, but since it is defined operationally by *E* manipulation, it is essentially a variable of *task structure* independent of *S*. Several studies have regulated task difficulty and/or intratask competition by manipulating the association values of the syllables involved. For example, Montague (1953) employed lists of different association values as one technique for varying intralist interference; Saltz and Hoehn (1957) controlled task difficulty by means of association values while varying intratask competition through manipulation of intralist similarity; and Sarason (1957) altered task difficulty by means of association values in testing the relationship between task difficulty and MAS level. In studies of this nature, it is implicitly assumed that the degree of task difficulty and/or competition is independent of anxiety drive as measured by the MAS. That is, the association values of the syllables are considered comparable for anxious and nonanxious groups at the beginning of the learning trials.

The present study is designed to test the independence of nonsense syllable association with respect to drive level as evaluated by MAS. Since both association to nonsense syllables and score on the MAS are *S* variables, covariance may possibly be found between anxiety level and association value. Davids and Eriksen (1955), employing a word association test, did find a significant positive correlation between MAS scores and productivity of associations, and Kausler and Trapp (1959) stated more explicitly the rationale and implications for this suspected relationship. If a significant relationship can be demonstrated between MAS scores and association values of nonsense syllables, there would be valid grounds for questioning the methodology in studies which control task properties in terms of association values of the individual items. For example, lists of syllables supposedly equated in difficulty for different levels on the MAS may actually represent lists of varying difficulty for different levels of the MAS.

---

[1] This study was supported by a grant from the College of Arts and Sciences Research Fund, University of Arkansas. Dale Wise assisted in collecting and analyzing the data.

METHOD

Subjects. The *S*s consisted of 90 male students, with approximately median ACE scores, in general psychology classes at the University of Arkansas. (Eight *S*s had to be dropped from the study for failure to follow the instructions.) Since some studies suggest that the MAS applied to an unselected population differentiates groups on sex and intelligence, the homogeneity of the present sample eliminated a potential bias. The *S*s were tested during regular class periods.

Association task. A list of 90 nonsense syllables was prepared for group administration in a booklet containing five pages with 18 syllables to the page, arranged in random order. The nonsense syllables were selected on the basis of meeting two criteria. The first requirement was that the association values reflect close agreement between the calibrations made by Hull (1933) and a revision of his work (Trapp & Kausler, 1959). The revision involved a group rather than an individual presentation of syllables. The aim was to choose nonsense syllables with highly stabilized association values, and this seemed best fulfilled by including nonsense syllables that had demonstrated relatively unchanged association values over a 25-yr. period. With very few exceptions, the selected nonsense syllables deviated less than 5% from the values obtained by Hull in spite of the different procedures used in administration. The second requirement was that the list of nonsense syllables contain equal proportions of syllables with low, intermediate, and high association values. The purpose of this requirement was to make possible an analysis of interaction between anxiety level and level of association value. Hence, 30 of the nonsense syllables had association values ranging from 5% to 15% (low values), 30 had values ranging from 20% to 30% (intermediate values), and 30 had values ranging from 35% to 66% (high values).

Procedure. The *S*s were informed that they were participating in a study of the association value of nonsense syllables. They were instructed to look at each syllable briefly, and if the syllable aroused any particular thought or idea, they were to write it down on the separate answer sheet. The use of a template restricted their attention to one syllable at a time.

Approximately two weeks later, *S*s were given the Taylor MAS, disguised as a biographical inventory, by a different *E*. The form used consisted of the 50 anxiety-keyed items and 40 buffer items. The *S*s were informed that *E* was collecting biographical information about students at Arkansas to compare with students elsewhere. Apparently none of the *S*s suspected any relationship between the two parts of the experiment.

RESULTS

For each *S* two sets of measurements were obtained: an MAS score as determined by the Taylor scoring system, and association scores (number of syllables responded to with an association) for the low, intermediate, and high levels of the list, and for the total list. The summary data showing the relationship between MAS score and association scores are given in Table 1. The MAS scores ranged from 3 to 32 with a median of 16. Total association scores for all 90 syllables ranged from 3 to 84 with a median of 41.5 and mean of 43.38. In order to determine the over-all degree of relationship between MAS score and total association score, a correlation ratio was computed between these two variables. The correlation ratio was used since the data suggested a curvilinear relationship. An eta of .41 ($P < .05$) was obtained, which was mainly attributable to the high association scores by *S*s in the upper range on the MAS.

TABLE 1

SUMMARY DATA ON LIST OF NONSENSE SYLLABLES (90 ITEMS) WITH THREE LEVELS OF ASSOCIATION VALUE FOR SCORES ON MAS

MAS Score	N	Association Value						All Syllables	
		Low		Intermediate		High			
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
0–9	11	7.82	8.56	13.09	8.56	18.27	8.99	39.18	24.90
10–11	12	8.00	8.28	12.75	10.39	18.83	9.54	39.58	26.82
12–13	9	8.78	9.87	14.33	11.29	19.89	11.95	43.00	31.94
14–16	10	9.20	8.36	12.40	9.17	18.70	10.57	40.30	27.13
17–18	9	3.67	3.49	8.56	6.72	16.89	9.26	29.11	17.01
19–20	9	15.22	9.68	19.11	11.96	25.22	11.15	59.56	31.05
21–24	10	7.40	9.93	11.50	10.26	17.00	10.61	35.90	29.77
25 & above	12	14.50	10.21	18.83	10.53	25.33	9.77	58.67	29.19
All *Ss*	82	9.40	8.66	13.90	9.17	20.07	8.35	43.38	24.30

Following the practice used in most studies with the MAS, *Ss* were assigned to one of three groups on the basis of their MAS scores. The 31 *Ss* scoring 19 or above constituted the high anxiety group (HA); the 19 *Ss* scoring between 13 and 19 constituted the medium anxiety group (MA); and the 32 *Ss* scoring 13 or below comprised the low anxiety group (LA). The MAS cutting scores were arbitrarily selected, representing the most logical and, considering ties, practical tripartite division of the total distribution of scores. As is apparent from Table 1, the various subgroups did not differ significantly in variability. Differences between the three levels of anxiety on the three levels of association values were then tested by an analysis of variance. The results of the analysis of variance are given in Table 2. Anxiety level was statistically significant ($F = 3.26$, $P < .05$), but the interaction between anxiety level and level of association value was not ($F = .55$). Hence, the effect of anxiety was independent of the level of association value.

Separate t tests were made between means for the three anxiety groups on the total list of syllables. The only significant difference was between Groups HA and MA ($t = 2.32$, $P < .03$, two-tailed test). The t values between Groups HA and LA and between MA and LA were 1.66 and .80, respectively.

The foregoing analyses were made in terms of differences between *Ss* on association scores. A different, and equally important, approach to the problem consists of analyzing the syllables themselves for differences in

TABLE 2

ANALYSIS OF VARIANCE WITH SUBJECT GROUPING FOR DIFFERENCES IN ASSOCIATION VALUES

Source	*df*	*MS*	*F*
Anxiety (A)	2	617.00	3.26*
Ss in same group	79	189.42	
Association value (AV)	2	2353.30	225.41**
A × AV	4	5.79	.55
AV × Pooled *Ss*	158	10.44	
Total	245		

* $P < .05$.
** $P < .01$.

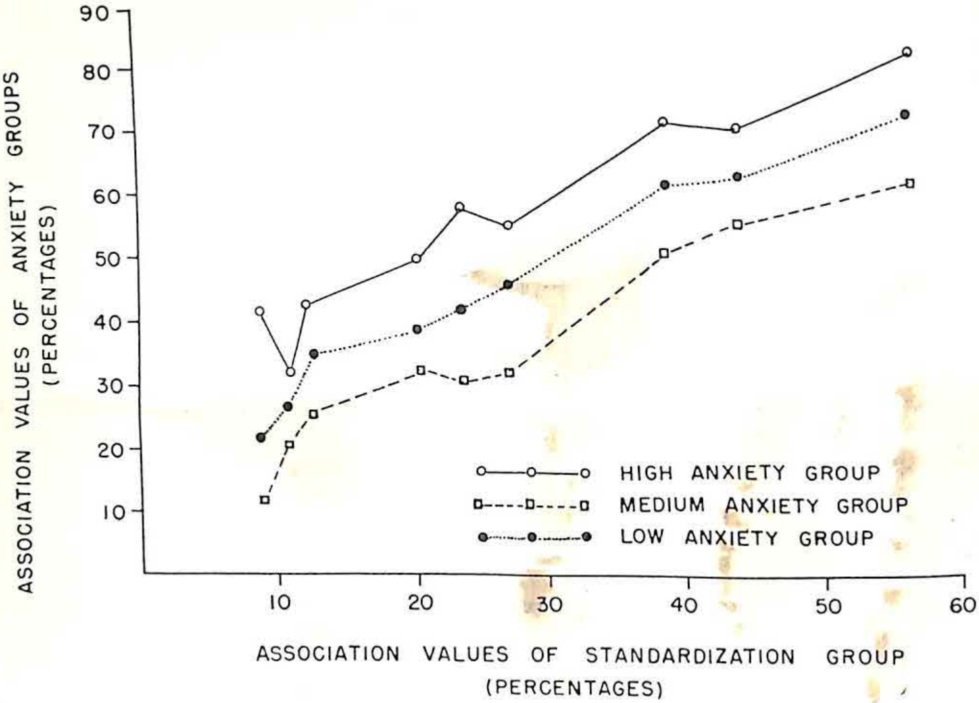


FIG. 1. Association values for groups differentiated on MAS.

association value as a function of anxiety level. For this analysis three association values, calibrated independently for each of the three anxiety groups (HA, MA, and LA) were determined for each of the 90 nonsense syllables. To simplify the analysis, the 90 syllables were grouped in unit blocks of 10 syllables per block and arranged in progressive order of their association values as calibrated from the preliminary standardization study. For each level on the MAS the average association value (percentage of response) was determined for each of the nine blocks of syllables. In Fig. 1 the association values are presented graphically for each MAS group at each point along the continuum representing association value as determined, independently of anxiety level, in the preliminary standardization study.

TABLE 3

ANALYSIS OF VARIANCE WITH SYLLABLE GROUPING FOR DIFFERENCES IN TRANSFORMED ASSOCIATION VALUES

Source	*df*	*MS*	*F*
Anxiety (A)	2	387.63	16.87*
Syllables (blocks of 10) in same group (S)	6	22.98	
Association value (AV)	2	1104.72	150.51*
A × AV	4	2.11	.29
AV × Pooled syllables	12	7.34	
Total	26		

* $P < .01$.

As can be readily observed from an inspection of Fig. 1, anxiety level influences association values at all points on the continuum. The effect is consistently in the same direction with Groups HA, LA, and MA ranking in that order at levels of association value.

In order to test for significant differences between groups, the percentages were transformed to arc-sines (following the procedure outlined by Snedecor, 1946) and treated by analysis of variance. The summary of this analysis is presented in Table 3. The *F* of 16.87 ($P < .01$) permits rejection of the hypothesis that no relationship exists between MAS level and association value of nonsense syllables. Again, no significant interaction effect was found between level of anxiety and level of association value ($F = .29$, $df = 4$ and 12). Separate *t* tests were again made between anxiety groups. The only significant bifference was detween Groups HA and MA ($t = 2.81$, $P < .02$). The *t* values between Groups HA and LA and between Groups MA and LA were 1.41 and 1.32, respectively.

DISCUSSION

The data, whether approached from the standpoint of *S*s' productions or from a direct analysis of the nonsense syllables themselves, indicate that association values of nonsense syllables are a function of anxiety level as evaluated by MAS. Lists of nonsense syllables serially presented, therefore, cannot be assumed to be of equal difficulty at the start of learning trials for groups differentiated on MAS if task difficulty and/or intratask competition is a function of the level of association value. Studies making this assumption have left an important variable uncontrolled, which makes it impossible to infer if the diffential performances obtained between MAS groups reflect differences in drive level, level of task difficulty or intratask competition, or some interaction effect.

From our results, it is apparent that this confounding interaction between MAS level and association value is most pronounced in comparisons between high anxious and moderately anxious *S*s. Although most studies in this general area ignore the middle anxiety range, an important research goal, nevertheless, is to extend the relationship of anxiety to human verbal learning throughout the entire anxiety continuum. In this extension, an important methodological consideration should be to control for the difference in association value within this range.

The nature of the relationship between association values and MAS poses an interesting theoretical problem and is deserving of some consideration even though it is incidental to the primary purpose of the study. In Fig. 1, it may be recalled, the highest association values were for Group HA, the next highest values for Group LA, and the lowest values for Group MA. The authors found a similar pattern in an independent preliminary study which strongly discounts the possibility of attributing this finding to an artifact of the experimental design. One possible interpretation—accepting for present purposes that MAS is measuring both general drive (Taylor, 1956) and levels of interfering or competing responses (Mandler & Sarason, 1952)—may be advanced as follows: MAS measured drive follows a positively accelerated curve whereas the number or strength of interfering responses project a negatively accelerated curve. Performance, facilitated by drive and inhibited by interfering responses, is conceived as a function of the net differential between the two factors. The largest differential would be among *S*s in Group HA (drive level having risen sharply but interference level having tended to flatten out) so their performance would be highest. On the other hand, *S*s in Group MA with only slightly higher drive level than *S*s in Group LA but with relatively many more interfering responses, would produce the smallest differential and, consequently, reflect lower performance than *S*s in Group LA.

One other point needs some comment. From Fig. 1 it may be noted that the experimental group as a whole (combining all levels of anxiety) reflected higher association values than the standardization group. This discrepancy can be readily explained in terms of differences in the task situation for the two groups. The standardization *S*s were presented with a list of 320 nonsense syllables whereas the experimental *S*s were presented with only 90 syllables from this list. It would be expected that fewer syllables would elicit a higher percentage of responses.

SUMMARY

The present study investigated the relationship between MAS scores and association values of nonsense syllables. Three groups of male *S*s (HA, MA, LA) were given 90 nonsense syllables, ranging from low to high association values. A correlation ratio was computed between MAS scores and number of associations, with an obtained eta of .41 ($P < .05$). The association values were grouped in three levels, and differences between the three levels of anxiety on these three levels of association values were tested by an analysis of variance. Anxiety level was statistically significant ($F = 3.26$; $P < .05$) but interaction was insignificant ($F = .55$). Separate t tests were additionally made, with the only significant difference found between Groups HA and MA ($t = 2.32$; $P < .03$). The nonsense syllables were also analyzed for differences in association value as a function of anxiety level. Three association values on blocks of 10 syllables were calibrated independently for each of the three anxiety groups, with significant differences obtained ($F = 16.87$; $P < .01$). A significant t value was found between Groups HA and MA ($t = 2.81$; $P < .02$). Again, no interaction was found. The methodological implications of the general findings were discussed.

REFERENCES

DAVIDS, A., & ERIKSEN, C. W. The relation of manifest anxiety to association productivity and intellectual attainment. *J. consult. Psychol.*, 1955, **19**, 219–222.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

HULL, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

KAUSLER, D. H., & TRAPP, E. P. Methodological considerations in the construct validation of drive-oriented scales. *Psychol. Bull.*, 1959, **56**, 152–157.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.

SALTZ, E., & HOEHN, A. J. A test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.

SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.

SILVERMAN, R. E., & BLITZ, B. Learning and two kinds of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 301–303.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

TAYLOR, J. A., & CHAPMAN, J. P. Paired-associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.

TRAPP, E. P., & KAUSLER, D. H. Revision of Hull's Table of Association Values for 320 selected nonsense syllables. *Amer. J. Psychol.*, 1959, **72**, 423–429.

(Received May 6, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

LATENCY OF INSTRUMENTAL RESPONSES AS A FUNCTION OF COMPATIBILITY WITH THE MEANING OF ELICITING VERBAL SIGNS[1]

ANDREW K. SOLARZ[2]

University of California, Davis

Does the ease of association of an instrumental movement with a verbal stimulus depend upon the compatibility of the movement with the meaning of that verbal stimulus? Such a relation would appear to follow from the two-stage mediational model described by Osgood (1953). He assumes that in the development of meanings of signs, either perceptual or verbal, the sign-stimuli become associated with distinctive portions of the total behavior elicited by significate-stimuli. This portion of the total behavior to things signified is called a *representational mediation process*. It is representational because it is part of the same behavior made to the significate; it is mediational because, through its self-stimulational properties, it can become associated with overt instrumental movements.

Among the classes of responses elicited by both signs and significates are *evaluative reactions*—approaching, avoiding, reaching for, pushing away, and so forth. If the total behavior elicited by a positive or negative significate includes such approach or avoidance movements, then the theory requires that tendencies or dispositions toward such movements must become part of the total mediation process (i.e., part of the meaning of the sign). Furthermore, since the perceptual and vocal signs of positively evaluated objects will often mediate approach movements and those of negatively evaluated objects avoidance movements, one may expect such mediators to tend to elicit overt approach or avoidance movements directly. There are, therefore, two bases in theory to expect easier learning of compatible signs and movements: the evaluative nature of the representational process itself (based on approach and avoidance movements originally made to the things signified) and the generalized association of such mediators with overt instrumental movements.

The purpose of the present study is two-fold: (*a*) to provide a situation in which a two-stage, mediation model is required to account adequately for the data; and (*b*) to provide some evidence for the essentially behavioral nature of mediation processes. A situation is required in which the verbal signs to be used as stimuli have not been previously associated with the particular approach and avoidance movements—otherwise no two-stage interpretation would be necessary. In the present experiment it is assumed that people have seldom, if ever, learned to pull *printed words* toward themselves or to push them away, even though they may have done so in connection with the things signified. If it can be shown experimentally that it is easier to learn to pull favorably evaluated words (like Fragrant, Kind, and Clean) toward oneself than to push them away, and conversely for unfavorably evaluated words (like Foul, Cruel, and Dirty), then this will be interpreted as evidence for a two-stage mediation model (since it is the meanings of the printed signs, not their physical stimulus properties, that are compatible or incompatible with the overt movements). To differentiate between the two alternative interpretations—compatibility of representational processes or of generalized overt movement tendencies with the new movements in the experimental situation—two response indicators are required: one for the time required to *initiate* the correct movement and the other for the time to *execute* the correct movement. If the differences between compatible and incompatible conditions are found to be restricted to the time to *initiate* the correct movements, then this will be interpreted as evidence that the representational mediation processes themselves include the evaluative characteristics of their behavioral origins.

---

[1] The study was conducted at the University of Illinois.

[2] The author wishes to thank C. E. Osgood for his many helpful suggestions on the design and preparation of this research. Hans Schmidt deserves special thanks for his collaboration on the statistical analysis of the data.

METHOD

Subjects. The *S*s were 20 male and 20 female undergraduate students from psychology courses at the University of Illinois. They were chosen on a voluntary basis with an inducement of 50 cents for participation.

TABLE 1

STIMULUS WORDS USED AS SIGNS

List A	List B	List C	List D	List E
1 sweet	tasty	fragrant	pleasant	fresh
1′ sour	putrid	foul	painful	stale
2 kind	sane	smart	patriot	honest
2′ cruel	insane	stupid	traitor	dishonest
3 valuable	good	nice	clean	pretty
3′ worthless	bad	awful	dirty	ugly
4 near	here	toward	this	close
4′ far	there	away	that	distant
5 happy	gay	joyful	healthy	merry
5′ unhappy	sad	mournful	sick	weary

Apparatus. The apparatus consisted of a movable stage mounted on a runway perpendicular to the ventral axis of *S*. Stimulus cards were inserted behind a vertically sliding door on the stage. The *S* manipulated the door of the stage with a hand lever on the handle base of the stage. As *S* pressed the lever, the door dropped exposing the stimulus card. The stage could be moved 7 in. forward or backward from its center position. White markers on the base of the stage and along the runners allowed for recentralization of the stage.

As the stage door dropped it completed two circuits, starting two .01-sec. Standard Electric timers. Moving the stage ¼ in. off its center position broke one timer circuit and stopped one of the clocks. The second timer continued until the base of the stage made contact with either end of the runway, whereupon the second circuit was broken. In this way two latencies were obtained, one consisting of the time between exposure of the stimulus card and initiation of arm movement, and the other the time for the response movement itself.

Design. Each *S* was told that the experiment was a study of reaction time. A list of 10 words was handed to *S* and he was told, "Look at this list of words and think about the meaning of each word in each pair of words." Three minutes were allowed for this preliminary task. After *S* had been shown how to manipulate the apparatus and told that the listed words would now appear on cards, the following instructions were completed:

> When I say 'ready' I want you to let the door drop, read the word on the card and then move it either toward or away from yourself. After your move I will say either 'wrong' or 'right' according to a prearranged pattern of correct moves for the various words. You are to try to learn which way to move each word as soon as you can and to react with that movement as quickly as you can. The move for each word is always the same.

Five lists were used (see Table 1). Each list contained 10 words, five pairs, which are presumed to be on an evaluative continuum. Each pair contained a positively evaluative word and a negatively evaluative word. The fourth pair in each list was included in order to accentuate the compatibility or incompatibility of the sign-movement relationship.

The response required to each word on each list was as follows for the original order of presentation: for the first word, evaluative sign—*pleasant*, the required movement was *toward* the body; for the second word, evaluative sign—*unpleasant*, the required movement was *away* from the body. The required movements for the third and fifth pairs were the same as above; however, the movements required for the second and fourth pairs were opposite, that is, an *away* movement to a *pleasant* word and a *toward* movement to an *unpleasant* word.

Two orders of response presentation were employed. In the reversed order the response movement required to each word was opposite to that in the original order. Consequently, under the original order, six of the sign-movement relation tasks were compatible and four were incompatible; the numbers interchanged under the reversed order. Each *S* under either order of presentation was thus required to learn to make five *toward* and five *away* movements to the words on the given list.

Twenty *Ss* served under each of the two orders of response presentation. Four *Ss* were used for each list in each order of presentation; they consisted of two men and two women. Each *S* was exposed to the words of only one list throughout the experiment.

A trial consisted of one complete presentation of all 10 stimulus cards. The order of word presentation for any list was varied from trial to trial in accordance with a 12 × 12 latin square so that each word appeared in a different sequence on each trial. This prevented *S*'s anticipating the next stimulus word.

The reason *S* was not informed of the criterion, two errorless trials, was to prevent his sacrificing speed of response for correctness of response; *S* was informed, however, that the time required for participation would not exceed 1 hr.

RESULTS

Mean latencies were obtained for each *S* for each of the four sign-movement relationships. The latencies for the first trial were not used since the responses here served merely to inform *S* which direction of arm movement was correct for the particular stimulus word. Only latencies of correct responses were used. Finally, these mean values were entered into grand means for the given conditions of the experiment.

TABLE 2

NUMBER OF LATENCY VALUES ENTERED INTO THE FINAL MEANS USED FOR ANALYSIS OF VARIANCE

Presentation	*N*	Sign-Movement Category[a]			
		++	−+	−−	+−
Original	20	326	203	329	189
Reversed	20	187	267	199	266

[a] First symbol designates sign; second symbol, movement.

Since the errors occurring were not equally distributed over the four categories, the final means are not based upon equal numbers of values, as indicated in Table 2. Also, as noted previously, the number of values entered for compatible and incompatible tasks differed for original and reversed presentation.

If the number of values entered into the final *S* means of each category is a factor in determining the reaction time obtained, a significant interaction involving order of presentation would be expected. If it were the only factor, a reversal in the direction of difference between the final sign-movement means would be anticipated. The presentation of the results shows no such interaction or reversal.

The analysis of variance was applied to the two sets of mean latency scores of individual *Ss* for the four sign-movement categories. The results of the analysis are presented in Table 3. As indicated under Clock 1 (time to initiate arm movement) the

TABLE 3

ANALYSIS OF VARIANCE OF SUBJECT MEAN LATENCIES FOR EACH SIGN-MOVEMENT

Source of Variation	*df*	Clock 1 Time to Initiate Arm Movement		Clock 2 Time to Complete Arm Movement	
		MS	*F*	*MS*	*F*
List	4	1176.57		336.80	2.53
Sex	1	191.42		1998.69	15.02***
Order	1	2768.91		333.22	2.50
List × Sex	4	1709.91		454.74	3.42*
List × Order	4	3597.75	1.24	15.86	
Sex × Order	1	64.75		.42	
List × Sex × Order	4	2568.62		76.55	
*S*s in same group	20	2895.38		133.04	
Total between *S*s	39	2490.84		218.69	
Move	1	216.23	1.32	655.69	62.75***
Word	1	556.53	3.40	17.89	1.71
Move × Word	1	3717.17	22.69***	7.61	
Move × Word × Order	1	47.55		4.01	
Word × Move × List	4	915.88	5.59***	14.98	1.43
Word × Move × Sex	1	1328.28	8.11**	18.98	1.82
Word × Move × List × Order	4	427.45	2.61*	5.88	
Other higher interaction effects	47				
Pooled *S*s × treatments	60	163.81		10.45	
Total within *S*s	120	176.20		16.99	

* $P = .05$.
** $P = .01$.
*** $P = .001$.

Move × Word interaction is highly significant. Whether initiating an arm movement toward oneself has a shorter or longer latency following stimulus presentation than an arm movement away from oneself is dependent upon whether the sign has a pleasant or unpleasant meaning.

TABLE 4

CLOCK 1 MEANS IN SECONDS FOR MEN AND WOMEN FOR EACH SIGN-MOVEMENT CATEGORY

Move	Men		Women	
	Pleasant Word	Unpleasant Word	Pleasant Word	Unpleasant Word
Toward	.803	.865	.760	.965
Away	.871	.856	.931	.827

The Clock 1 means in seconds for all *S*s for each sign-movement category were: pleasant-toward = .781; pleasant-away = .901; unpleasant-away = .842; unpleasant-toward = .915. The time needed to initiate a movement to a sign is less when the meaning of the sign and the direction of the movement are compatible than when they are incompatible.

Table 3 also indicates a significant interaction of Word × Move × List for Clock 1. An inspection of the four sign-movement means for each list revealed that List D contributed the greatest total difference, .419 sec.; the shorter latencies being associated with compatible sign-movements. List C showed a reversal in the direction of the difference, a value of .163 sec. Lists A, B, and E corresponded to List D, showing a range of differences of .199 sec. to .272 sec. in favor of quicker compatible sign-movements.

The Word × Move × Sex interaction is also significant for Clock 1 in Table 3. Table 4 presents the sign-movement means separated for sex. Women contributed significantly more to the direction of the differences than did men.

Order of response presentation itself was not statistically significant nor was it a significant variable in the interaction effects of Move × Word, as indicated in Table 3, Clock 1. However, when the interaction effects of Move × Word are broken down into Lists, order of presentation becomes a significant interacting variable.

Turning attention to the results of the analysis applied to Clock 2 (time of the actual arm movement itself) in Table 3, it may be seen that sex is a significant variable for response time. The mean response time for men was .324 sec., for women .394 sec. The significance of the List × Sex interaction indicates that the speed of response for men and women was differentially affected by the particular list of words employed.

There is no significant relation between sign meaning and direction of movement once the response has been initiated, as indicated by the Move × Word interaction.

The last significant variable in the Clock 2 analysis is the direction of arm movement for all *S*s. The mean time for arm movement toward the body is .339 sec. and for movement away from the body .379 sec. This is arm movement from the center of the apparatus for a distance of 7 in. toward either the near or far end of the apparatus.

An error-difference score was computed for each *S*. This was obtained by subtracting his mean error on compatible sign-movement from that on incompatible sign-movement. For the group, the mean error-difference score was .468 ($t = 3.72$, $df = 39$, $P < .01$). Separating for sex, the mean error-difference scores were, men .629 ($t = 3.14$, $df = 19$, $P < .01$), women .307 ($t = 2.07$, $df = 19$, $P > .05$). The direction of the differences was in favor of fewer errors on compatible tasks.

The final analysis consisted of a comparison of the combined error scores for men (mean error score on compatible tasks *plus* that on incompatible tasks) with the combined error scores for women. The difference between the means (men 1.222, and women .829) was not significant ($t = 1.81$, $P > .05$).

DISCUSSION

Recent studies (Fitts & Deininger, 1954; Fitts & Seeger, 1953) have indicated shorter latencies and fewer errors when the stimulus of a given stimulus set was spatially compatible with the required response. Stimuli consisted of spatial arrangements of lights or symbols. The results of the present study point out that compatibility extends beyond objective S-R spatial congruity to the subjective level of the compatibility of the sign-meaning with the response-meaning.

In terms of Osgood's (1953) two-stage mediational model of behavior, the learning of movements incompatible to signs would involve some degree of change, either temporary or permanent, in the representational mediation process associated with the stimulus; that is, the sign would assume a postexperimental shift in meaning, particularly within the experimental setting. When the signs and movements are compatible, learning would consist of a strengthening of the already existing association between sign and mediation process, and a strengthening of the required response within the already associated hierarchy of responses. In this connection note the finding of Solley, Jackson, and Messick (1957) that *S*s, when asked to anticipate, made more guesses for cards containing profiles which had previously been moved "toward" them than for the cards containing profiles without this prior "toward" association. Their finding, as the present one, suggests a perceptual relationship which may serve as reinforcement particularly as regards mediation processes.

The finding of shorter latencies and fewer errors for compatible sign-movements suggests that *S*s came into the experiment with established mediation processes which could be evoked by the proper signs. The existing mediation process associated with a particular evaluative sign was evoked by that sign regardless of the compatibility or incompatibility of the required movement. Osgood states, "Stimulus patterns associated with the same stimulus objects will acquire common mediating reactions and hence will become signs having similar significance" (Osgood, 1953, p. 402). Any particular mediation process is associated with a hierarchy of overt responses, the selection of which is partially determined by the immediate demands of the environment. It is assumed that the hierarchy available to a mediation process, evoked by pleasant or unpleasant signs, contains either responses serving to decrease the distance (sign pleasant) or increase the distance (sign unpleasant) between oneself and the given sign.

Attention is called to the fact that performing incompatible sign-movements did not involve the time to execute the instrumental response (arm movement) itself. One would expect on the basis of animal experiments (Brown, 1948; Bugelski & Miller, 1938) that the rate and strength of an overt response would be affected when an approach or avoidance movement is made to an incompatible evaluative sign. Hovland and Sears (1938) found "blocking and compromise responses" predominating as modes of resolution to approach-avoidance and avoidance-avoidance conflicts. If overt movements contained the evaluative characteristics due to past associations, then the generalized overt movement tendencies to incompatible signs should have resulted in longer times of execution. This was not the case. The results support Osgood's view that the representational mediation processes include the evaluative characteristics of their behavioral origins.

SUMMARY

Twenty men and 20 women served in a study of the relation between arm movements, toward or away from the body, and positively or negatively evaluative signs (words like Smart, Stupid, Happy, etc.). Printed word cards were presented on a movable stage and *S* was required to learn to move the signs according to a prearranged assignment of signs to compatible or incompatible movements. Five lists of 10 words each were employed in the experimental design.

Results revealed shorter latencies of initiating a "toward" or "away" movement to a compatible sign than to an incompatible sign. The compatibility variable did not affect the time of execution of the arm movement itself. In addition, *S*s committed more errors in learning incompatible tasks than compatible ones. Differences between men and women in performance were found on latency measures and on errors of response. The results were interpreted in terms of Osgood's theory of "meaning as a representational mediation process."

REFERENCES

BROWN, J. S. Gradients of approach and avoidance responses and their relation to motivation. *J. comp. physiol. Psychol.*, 1948, **41**, 450–465.

BUGELSKI, R., & MILLER, N. E. A spatial gradient in the strength of avoidance responses. *J. exp. Psychol.*, 1938, **23**, 494–505.

FITTS, P. M., & DEININGER, R. L. S-R compatability: Correspondence among paired elements within stimulus and response codes. *J. exp. Psychol.*, 1954, **48,** 483–492.

FITTS, P. M., & SEEGER, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46,** 199–210.

HOVLAND, C. I., & SEARS, R. R. Experiments on motor conflict and their modes of resolution. *J. exp. Psychol.*, 1938, **23,** 477–493.

OSGOOD, C. E. *Method and theory in experimental psychology*, New York: Oxford Univer. Press, 1953.

SOLLEY, C. M., JACKSON, D. N., & MESSICK, S. J. Guessing behavior and autism. *J. abnorm. soc. Psychol.*, 1957, **54,** 32–36.

(Received May 11, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

EFFECT OF DISCRIMINATION TRAINING ON AUDITORY GENERALIZATION[1]

HERBERT M. JENKINS

Bell Telephone Laboratories, Incorporated

AND ROBERT H. HARRISON

Massachusetts Mental Health Center

Two types of training procedures have been used in studies of stimulus generalization. In the nondifferential procedure a response is simply reinforced repeatedly in the presence of, or closely following the presentation of, a stimulus of a particular value. In the differential procedure, on the other hand, a discrimination is developed between two stimulus values by reinforcing the response to one value and extinguishing the response when it occurs to the other value.

Lashley and Wade (1946) argued that differential training was required for the development of a gradient of generalization. The several experiments which faced this issue (Grandine & Harlow, 1948; Grant & Schiller, 1953; Grice, 1948) have demonstrated, however, that a sloping gradient of generalization can be obtained following a nondifferential training procedure.

In experiments comparing gradients of generalization following training with these two procedures (Hovland, 1937a; Schlosberg & Solomon, 1943); it was found that differential training produced a steeper gradient. Although the differential training involved nonreinforcement with respect to a stimulus that differed from the reinforced stimulus along the dimension on which the gradient was obtained, it is not clear that the effect depends entirely on this feature. The purpose of the present experiments was to examine the possibility that differential training not involving an explicit discrimination of this kind also produces a steeper gradient.

The method was similar to the one used by Guttman and Kalish (1956). The key-response of the pigeon was reinforced in the presence of a pure tone. Generalization gradients along the dimension of frequency were compared following differential or nondifferential training. Differential training was used to produce a discrimination between the presence of tone and its absence; it did not involve explicit training to discriminate frequencies. The expectation that training of this form sharpens the gradient of generalization was based on the following reasoning. The establishment of a discrimination between presence and absence ensures that the tone has discriminative control over the response. This provides the potential for an increase in discriminative control by any of the several properties of the tonal stimulus; properties which might be expected to include frequency as well as intensity, rate of interruption, and so on.

---

[1] This research was done at the Massachusetts Institute of Technology and was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract AF 18 (603)-85. The present article is a condensed version of the final report on that contract; TN 58-443, ASTIA Document 158248. Reproduction for any purpose of the United States Government is permitted.

The authors wish to express their appreciation to Carl Hovland for his helpful criticisms of a draft of this article.

EXPERIMENT I

Method

Subjects. The *S*s were eight male, white Carneau pigeons maintained by restricted feeding at 80% of their body weight under free feeding.

Apparatus. A Skinner automatic key-pecking apparatus was used. The general features of the apparatus have been described elsewhere (Ferster & Skinner, 1957). The reinforcement was a 4-sec. period of access to a tray of mixed grain. The occurrence of a reinforcement was signaled by turning off the illumination of the response-key, and by turning on illumination for the food tray.

The auditory stimuli were pure tones sounded by a 5-in. speaker suspended 3 in. behind the response-key. The tone was interrupted for 0.25 sec. in each second in order to reduce possible adaptation to a steady tone. The amplification of tones at different frequencies was adjusted to give a sound pressure level of 70 decibels re 0.0002 dynes/cm.^{2} as measured with the aid of a calibrated microphone.

Preliminary training. Four sessions of preliminary training were given. The key-response was conditioned using a method of successive approximations and then placed on a variable interval (VI) schedule of reinforcement. The tone which served as the reinforced stimulus, or S^D, was presented as soon as several responses occurred. The third and fourth sessions were divided into 25 working periods of 33-sec. duration by the introduction of 7-sec. blackout intervals. In each working period, the S^D was presented and responses were reinforced on the VI schedule which delivered from 1 to 3 reinforcements with an average of 1.7 reinforcements.

Procedure. In the nondifferential training method, sessions consisted of 25 working periods in which a tone at 1000 cps (S^D) was presented and responses were reinforced on the VI schedule. Three *S*s received 10, 20, or 40 daily sessions of nondifferential training.

In the differential training sessions, the working periods were of two types: S^D periods as in nondifferential training, and S^Δ periods in which no tone was presented and responses were nonreinforced. Training sessions consisted of 25 S^D periods, and a number of S^Δ periods which was increased from 25 in the first session to 125 in later sessions. The S^D and S^Δ periods were presented in a random order. Daily training sessions were continued until the average rate of response in the S^D periods was at least four times greater than in S^Δ periods in each of two successive sessions. When testing the discrimination against this criterion, the sessions consisted of 25 S^D periods and 50 S^Δ periods. Five *S*s received differential training. From six to nine sessions were required to reach the criterion of discrimination.

In testing for generalization, the stimuli were presented for 33-sec. periods separated by 7-sec. blackouts as in training. Eight test stimuli were used. They consisted of seven tones approximately equally spaced along a logarithmic scale of frequency. Their frequencies were 300 cps, 450 cps, 670 cps, 1500 cps, 2250 cps, 3500 cps, and 1000 cps (S^D in training). The eighth stimulus was no tone. In a test session, the eight stimuli were presented in eight blocks, making a total of 64 presentations. The order of presentation was governed by 8 × 8 latin squares so that each stimulus appeared once in every ordinal position in the block.

Tests for generalization were conducted without reinforcement except where noted. Three tests for generalization separated by single retraining sessions were usually made. The procedure during retraining was the same as that used during training.

Results

Nondifferential training. The percentage of responses made to each test stimulus following nondifferential training is shown in Fig. 1. Prior to the generalization tests, S_{70} received 10 training sessions, S_{71} received 20,

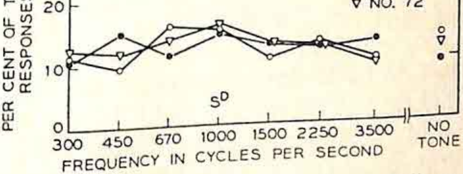


FIG. 1. Generalization gradients following nondifferential training with a 1000-cps tone as S^D. Individual gradients are based on the means of several generalization tests (see text).

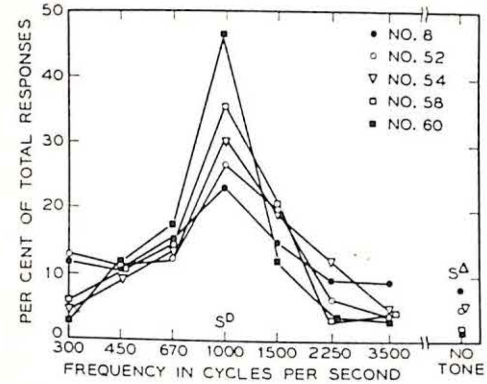


FIG. 2. Generalization gradients following differential training with a 1000-cps tone as S^D; no tone as S^{Δ}. Individual gradients are based on the means of three generalization tests.

and S_{72} received 40. Gradients are based on three tests on S_{70} and S_{71} and on two tests for S_{72}. Tones widely separated in frequency from the S^D, as well as no tone, were responded to in approximately equal measure. Increasing the amount of nondifferential training from 10 to 40 sessions did not have a marked effect upon the gradient.

There was some evidence for a very shallow gradient of response strength. An average of the individual curves would show the highest percentage of response at the S^D with a slight but systematic decrease at tones of higher or lower frequency.

A four-way analysis of variance was performed on the number of responses in 33-sec. test periods. The variables were *S*s confounded with amount of training, stimuli (eight values including no tone), tests (first and second generalization tests), and blocks within tests (eight blocks of eight stimulus presentations). The main effect of stimuli was significant at the .05 level ($F = 2.38$; 7/273 *df*). Since the simple interactions involving stimuli were not significant, the small main effect of stimuli provides some evidence for a shallow gradient following nondifferential training. Large and highly significant main effects were found for tests ($F = 229.54$; 1/273 *df*, $P < .001$) and for blocks within tests ($F = 36.36$; 7/273 *df*, $P < .001$). The effect due to tests reflects a sharp decline in the number of responses from the first to the second test. This was expected on the basis of the well established finding that extinction occurs more rapidly in successive extinctions following reconditioning (Bullock & Smith, 1953). The effect of blocks reflects a decline in the rate of response due to nonreinforcement during the tests of generalization.

Differential training. Generalization gradients following differential training are shown in Fig. 2. During the test sessions, S_8, S_{52}, and S_{54} received four additional presentations of the S^D in which one reinforcement was delivered. These presentations were excluded from the data. Differential training resulted in well defined gradients along the dimension of frequency. The percentage of responses to extreme frequencies (300 cps or 3500 cps) was only slightly greater than to no tone. The possibility that tones at these frequencies were inaudible, can be ruled out on the basis of data to be presented from Exp. II.

The analysis of variance for the differential group was based on S_{58} and S_{60} which received no reinforcement during the tests for generalization. Three tests for generalization were included in this and subsequent analyses. As is apparent from Fig. 2, the effect of stimuli was large ($F = 47.74$; 7/273 *df*, $P < .001$). The effects of blocks within tests ($F = 8.68$; 7/273 *df*, $P < .001$) and of tests ($F = 14.05$; 2/273 *df*, $P < .001$) again reflect, respectively, extinction within tests, and more rapid extinction in successive tests. Significant interactions were obtained for the effects of stimuli with blocks ($F = 1.80$; 49/273 *df*, $P < .01$), and of stimuli with tests ($F = 2.28$; 14/273 *df*, $P < .01$). The interaction effects were due in part to a progressive sharpening of the generalization gradient both from the beginning to the end of a test, and in successive tests. Similar trends are examined in more detail with the results of Exp. II.

EXPERIMENT II

Experiment II was designed to extend the comparison of generalization gradients obtained following differential or nondifferential training to a case in which responses are reinforced to tones at two frequencies (cf. Kalish & Guttman, 1957).

Method

The same method was used as in Exp. I except for the details noted below.

Procedure. Nondifferential training sessions consisted of 30 S^D periods. A 450-cps tone was presented in 15 of these periods; a 2500-cps tone was presented in the remaining 15 periods. The tones were presented in a random order. Three *S*s received 10, 20, or 40 daily sessions of nondifferential training.

The differential training procedure involved the same S^D periods, and in addition, S^Δ periods in which no tone was presented and responses were nonreinforced. The number of S^Δ periods was increased from 30 in the early training sessions to 150 in later sessions. This procedure does not involve explicit training to discriminate between the two tones used as an S^D.

Daily sessions were continued until the criterion of discrimination was reached (4 to 1 ratio of S^D to S^Δ rate of response) for each tone with respect to no tone. Six *S*s received differential training. From 7 to 11 training sessions were required to reach the criterion of discrimination.

The tests for generalization were made with the following stimuli: tones at frequencies of 250 cps, 450 cps (S^D), 670 cps, 1000 cps, 1500 cps, 2500 cps (S^D), 4000 cps, and no tone. Generalization tests were conducted without reinforcement. Three tests were made with a single retraining session between tests.

Results

Nondifferential training. The gradients obtained following nondiffertial training are shown in Fig. 3. Prior to the generalization tests, S_{29} received 10 training sessions, S_{30} received 20, and S_{36} received 40. The gradients are similar to those obtained following nondifferential training with respect to a single frequency in that they show a nearly equal percentage of response to all test stimuli, including no tone. Again, it is apparent that increasing the amount of training did not greatly alter the result. As in Exp. I, there was some evidence for a shallow gradient; each *S* had a local peak in response strength at 450 cps, although the curves over all test stimuli were irregular.

An analysis of variance for this condition shows effects similar to those for the nondifferential condition in Exp. I. The effect of stimuli was again significant ($F = 3.23$; 7/448 *df*, $P < .01$), and the simple interactions of stimuli with blocks or tests were not significant.

Differential training. Generalization gradients following differential training are shown in Fig. 4. Differ-

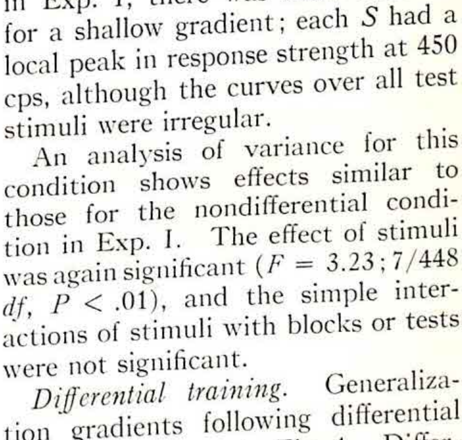


FIG. 3. Generalization gradients following nondifferential training with a 450-cps tone and a 2500-cps tone as S^Ds. Individual gradients are based on the means of three generalization tests.

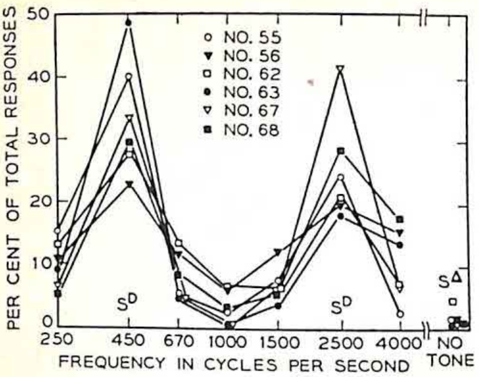


FIG. 4. Generalization gradients following differential training with a 450-cps tone and a 2500-cps tone as S^Ds; no tone as S^Δ. Individual gradients are based on the means of three generalization tests.

ential training resulted in much sharper gradients with peaks in response strength occurring at the frequencies reinforced in training. The effect of differential training in sharpening the gradient is consistent with the effect obtained in Exp. I.

The gradients with peaks near the extremes of the frequency range permit us to reject an interpretation of the low response strength obtained in Exp. I at the extreme frequencies of 300 cps and 3500 cps (see Fig. 2) as resulting from the inaudibility of these frequencies to the pigeon. It is also of interest to note that, in the present experiment, response strength at 1000 cps was only slightly greater than that for no tone; a stimulus which approximated silence. Although these observations do not eliminate the possibility that differences in loudness affected the gradients obtained, they make it appear unlikely that such differences were a major source of variance.

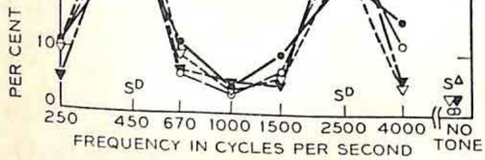


FIG. 5. Average generalization gradients for successive tests; Exp. II.

An analysis of variance shows effects that parallel those obtained with differential reinforcement in Exp. I with some exceptions. In particular, the interaction of *S*s with stimuli was highly significant ($F = 5.44$; 35/973 *df*, $P < .001$), whereas it was not significant in Exp. I. The major source of this interaction was in the difference among the *S*s in the S^D (450 cps or 2500 cps) at which they responded most frequently in the generalization tests. A change in the distribution of responses to the test stimuli both in successive blocks within tests and in successive tests was reflected in significant interactions of stimuli with blocks ($F = 2.07$; 49/973 *df*, $P < .01$) and of stimuli with tests ($F = 4.01$; 14/973 *df*, $P < .001$). These effects were also obtained in the differential group in Exp. I, and in both cases, represent

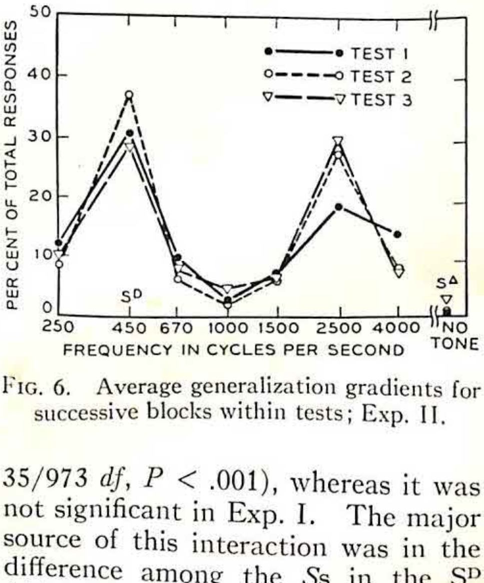


FIG. 6. Average generalization gradients for successive blocks within tests; Exp. II.

a general sharpening of the gradient from the first to later tests, and within the course of a single test. Figure 5 shows the mean percentage of responses to the test stimuli in successive tests. Figure 6 plots these data for successive blocks within the tests.

DISCUSSION

The experiments provide clear evidence that differential training can greatly sharpen a gradient of generalization even when this training does not involve an explicit discrimination between stimulus values along the dimension of the gradient. The conclusion stated by Lashley and Wade (1946, p. 74): "the 'dimensions' of a stimulus series are determined by a comparison of two or more stimuli and do not exist for the organism until established by differential training" is weakened by this result if the phrase "two or more stimuli" is taken to mean stimuli having values on the dimension of the gradient. It now seems likely that the absence or near absence of a sloping gradient along some dimension of a stimulus does not imply that training to discriminate values on this dimension is required. The present results suggest that the failure to obtain such a gradient may be due to a lack of control over the occurrence of the response by the experimental stimulus. When presenting or removing the stimulus has little or no effect upon the occurrence of the response, we can only conclude that the stimulus, and consequently any dimension of the stimulus, is irrelevant to the response.

The results pose a theoretical question: how account for the effectiveness of nonreinforcement in the absence of the tone in placing the tone in control of the response? A discussion of stimulus generalization by Hull (1952, p. 59 ff) is relevant here. He pointed out that in addition to the experimental stimulus, incidental stimuli (e.g., stimuli arising from constant features of the apparatus) accompany the reinforcement of the response. When training is nondifferential, these stimuli are repeatedly reinforced and may predominate in control of the response with the result that the gradient of generalization observed upon varying the experimental stimulus is flat or nearly so. When, on the other hand, training is differential, these incidental stimuli are both reinforced in the presence of S^D and nonreinforced in its absence. This reduces their effectiveness in controlling the response. Hull assumed, as does Restle (1955), that stimuli accompanying reinforced and nonreinforced responses adapt out, or in some manner become nonfunctional. The resulting increase in the relative weight of the experimental stimulus in controlling the response makes it possible to obtain a sharply sloping gradient of generalization along some dimension of the experimental stimulus. It is of interest to note that, according to this account, differential training involving neighboring stimulus values would also weaken the control by incidental stimuli. The effect of that procedure in sharpening the gradient might be largely due to this factor rather than due to the presumed development of a counter gradient of conditioned inhibition centered around the value of the nonreinforced stimulus. In an experiment by Reinhold and Perkins (1955) it was found that differential training on a feature of a compound stimulus unrelated to the feature used in testing for generalization produced a gradient with a steeper slope than did nondifferential training. They suggested that a set to discriminate was developed by differential training and that this resulted in less generalization. However, their results could also be accounted for by the reduction in effectiveness of incidental stimuli through differential training.

A weakness in the account offered by Hull and Restle lies in the specification of the conditions for adaptation of incidental stimuli. It is not sufficient to state that a stimulus accompanying both the reinforcement and nonreinforcement of a response adapts out, since in the present experiment responses to the S^D were reinforced only intermittently, yet with differential training the S^D showed rather precise discriminative control over the response.

Regardless of the interpretation placed upon the present results, it is clear that a property of a stimulus that shows little or no influence on a response after prolonged nondifferential training may gain rather precise discriminative control by means of differential training to discriminate the presence from the absence of the stimulus.

The auditory generalization gradients may be compared with the visual generalization gradients along the dimension of wavelength obtained by Guttman and Kalish (1956) using pigeons. In contrast with the present results, these investigators found steep, well-defined gradients following nondifferential training. Since the training procedures appear to be the same in all important respects it may be concluded that the difference lies in the use of a visual—as compared with an auditory stimulus. It does not appear that a lack of physiological capacity to detect or to make discriminations among auditory stimuli can account for the contrast in the results obtained with visual and with auditory stimuli. The more important role of visual stimuli in the normal activities of pigeons may be a factor. In locating food or in flying, visual discriminations are more heavily involved.

It should be noted, however, that the visual stimulus appeared directly on the response key, whereas the auditory stimulus was diffuse and probably unlocalized. Nondifferential training with a visual stimulus presented in an analogous manner, as an over-all illuminant, might also result in a flat generalization gradient.

Auditory generalization gradients following differential training and visual generalization gradients following nondifferential training are similar in some respects. In each case the typical gradient is negatively accelerated and approximately symmetrical. However, the visual gradients were plotted on a linear scale of wavelength, while the auditory gradients were plotted on a logarithmic scale of frequency. The approximate symmetry of the auditory gradients when plotted in this way may be related to the fact that human discriminability of frequency changes is proportional to the logarithm of frequency for tones above 1000 cps.

The present results, showing that the generalization gradients became sharper in the course of testing, are at variance with those reported by Kalish and Guttman (1957), who found no systematic change in the relative frequencies of response to test stimuli in the course of extinction. Some investigators have found sharpening (Brown, 1942; Hovland, 1937b; Littman, 1949), as in the present case, while others have found that gradients flatten during extinction (Grant & Schiller, 1953; Reinhold & Perkins, 1955). The question of what determines how extinction affects a gradient of generalization is still open.

SUMMARY

An operant conditioning method was used to obtain auditory generalization gradients along the frequency dimension of pure tones for the pigeon. Two procedures for training prior to testing for generalization in extinction were compared. In the differential procedure responses were reinforced in the presence of a tone and nonreinforced in the absence of a tone so that an explicit discrimination between presence and absence was formed. In the nondifferential procedure responses were simply reinforced in the presence of a tone.

In Exp. I the S^D was a tone at 1000 cps. In Exp. II equal reinforcement was given with respect to two S^Ds: a tone at 450 cps and one at 2500 cps. Test stimuli were tones higher and lower in frequency than the S^D (or S^Ds) and the absence of a tone.

Gradients of generalization following nondifferential training were nearly flat; that is, tones of different frequency, as well as the absence of a tone, were responded to about equally. On the other hand, well-defined gradients of generalization with relatively steep slopes were obtained following differential training. When responses were reinforced in the presence of two tones of different frequency, the gradients showed two peaks in the strength of response, one at each of the frequencies reinforced in training. To a

first approximation, the gradients were symmetrical about the S^D or S^Ds when plotted on a logarithmic scale of frequency.

It is concluded that training to discriminate the presence from the absence of a stimulus can greatly enhance the amount of control exerted over a response by a specific property of the reinforced stimulus.

REFERENCES

Brown, J. S. Generalized approach and avoidance responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.

Bullock, D. H., & Smith, W. C. An effect of repeated conditioning-extinction upon operant strength. *J. exp. Psychol.*, 1953, **46**, 349–352.

Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

Grandine, L., & Harlow, H. F. Generalization of the characteristics of a single learned stimulus by monkeys. *J. comp. physiol. Psychol.*, 1948, **41**, 327–338.

Grant, D. A., & Schiller, J. J. Generalization of the conditioned galvanic skin responses to visual stimuli. *J. exp. Psychol.*, 1953, **46**, 309–313.

Grice, G. R. The acquisition of a visual discrimination habit following responses to a single stimulus. *J. exp. Psychol.*, 1948, **38**, 633–642.

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148. (a)

Hovland, C. I. The generalization of conditioned response. IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses. *J. exp. Psychol.*, 1937, **21**, 161–176. (b)

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Kalish, H. I., & Guttman, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.

Littman, R. A. Conditioned generalization of the galvanic skin reaction to tones. *J. exp. Psychol.*, 1949, **39**, 868–882.

Reinhold, D. B., & Perkins, C. C., Jr. Stimulus generalization following different methods of training. *J. exp. Psychol.*, 1955, **49**, 423–427.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

Schlosberg, H., & Solomon, R. L. Latency of response in a choice discrimination. *J. exp. Psychol.*, 1943, **33**, 22–39.

(Received May 11, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

EFFECTS OF EXTINCTION TRIALS ON DISCRIMINATION REVERSAL [1]

M. R. D'AMATO AND H. JAGODA

New York University

In recent years, considerable interest has been shown in discrimination reversal learning. One important feature of such learning is that, in general, reversal learning (i.e., the first reversal) takes longer than the learning of the original discrimination, at least in rats (Dufort, Guttman, & Kimble, 1954; Kendler & Lachman, 1958). The present research is addressed to this problem. Previous observations, and certain theoretical considerations, led us to suspect that reversal learning was, in part, retarded by the "cue" function of the reward present during the reversal trials. Quite often during reversal learning, animals were observed to revert to the original discrimination habit after receiving a reinforcement on a reversal choice. This behavior seemed theoretically reasonable from the standpoint of the r_g mechanism, or even from some notion as the "aftereffects of reward" as a stimulus factor. It appeared plausible that extinction of the original discrimination was impeded by this cue function of reward, and that, conversely, if reward were removed from the situation, at least for a time, reversal learning might be facilitated.

To test this hypothesis we ran two groups of *S*s on a brightness discrimination problem. Group R was reversed in the normal way upon reaching criterion on the original discrimination. The experimental group, Group ER, first underwent a series of extinction trials after meeting the discrimination criterion. That is to say, Group ER learned the discrimination, underwent 60 extinction trials, and then entered reversal learning. Actually, the extinction trials may be considered reversal learning trials without reward. Our hypothesis then was that the extinction trials would, owing to the elimination of reward from the situation, lead to a more rapid extinction of the original discrimination, and that upon reinstitution of reward reversal learning would therefore occur faster in Group ER than in Group R, for which all reversal trials were rewarded.

METHOD

Subjects. The *S*s were 30 experimentally naive male albino rats, 60–90 days of age.

Apparatus. The apparatus consisted of three units of an automatic learning-discrimination apparatus (ALDA) recently developed at New York University.[2] Essentially, ALDA consists of a symmetrical Y maze (Fig. 1), with remote programming, switching, and recording apparatus. The presently available discriminanda are restricted to variable intensities of flashing (90 flashes/min.) and steady illumination in the arms of the Y maze.

A trial proceeds as follows. Assume that a brightness discrimination problem has been

---

[1] This research was facilitated by Grant M-2051 from the National Institute of Mental Health.

[2] The senior author is responsible for the design of the apparatus described and the supervision of its construction, financial support arising from a grant from New York University. J. Wilson and R. Lachman contributed imaginatively, and thanks are due R. Berryman for originally suggesting the symmetrical Y maze. H. H. Kendler provided essential support and encouragement throughout the project.

A more complete description of the apparatus is available from the senior author upon request.

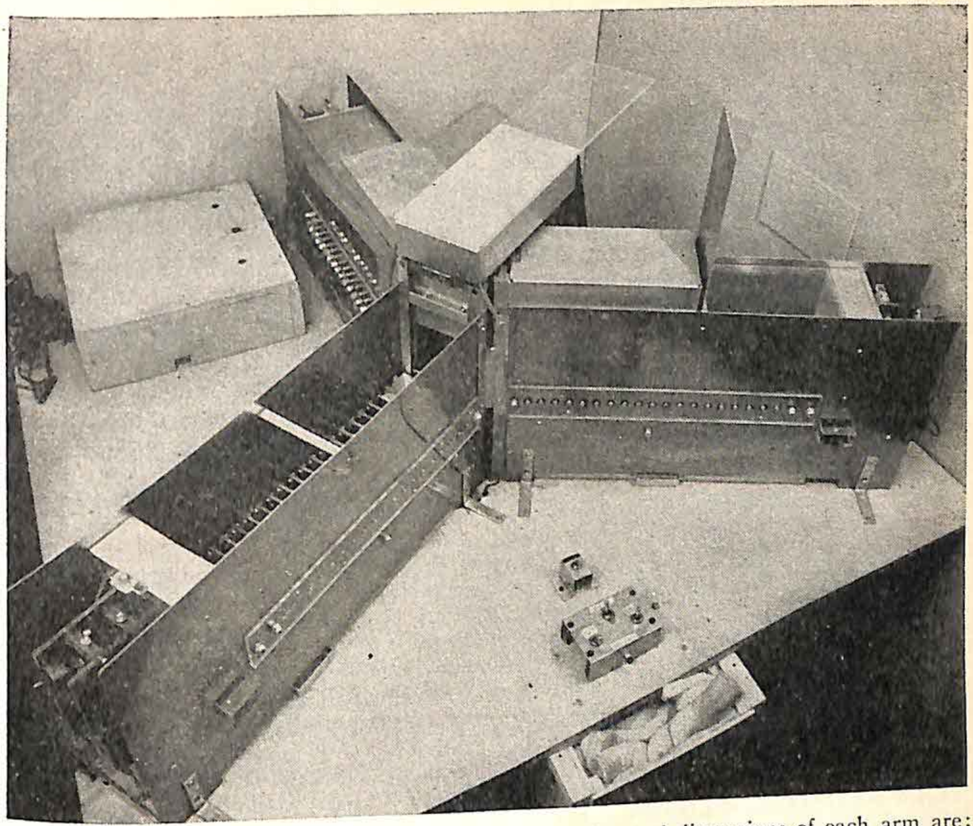


FIG. 1. The symmetrical Y maze of ALDA. The internal dimensions of each arm are: 17 in. long; 5 in. wide; 5.25 in. high. The discriminanda are presented through the translucent plastic of the lid and rear wall.

programmed, with S_+ being the brightly lighted arm and S_- the opposing dimly lighted arm. The *S* is placed into one of the arms, say Arm 1, and the starting button is pressed. The flashing light ("starting" signal) comes on in Arm 1, and at the same time S_+ occurs in one of the other arms, say Arm 2, and S_- appears in Arm 3. At the end of the starting interval (I_4), the doors open and *S* is confronted with the two differently illuminated arms.

Assume *S* makes a correct choice, i.e., he enters the brightly lighted arm. As soon as he penetrates the arm beyond a certain point, a low amperage transistorized contact relay activates, resulting in the closing of the doors and the stopping of the running-time timing mechanism. (Running time is measured from the opening of the doors until their closing.) However, nothing further happens until *S* starts taking his reward (water, sucrose solution, etc.). The first lap of the reward starts the reward interval (I_1) timing, at the end of which the reward dipper retracts out of sight. As soon as I_1 is finished the "stimulus interval" (I_2) starts timing. The purpose of I_2 is to provide a controllable period during which the discriminanda are present but reward is absent. I_3, the "intertrial interval," follows upon I_2. During the intertrial interval the illumination in the maze may be changed in accordance with one's requirements. For example, in the present study all lights were extinguished during the intertrial interval. At the end of I_3, I_4 (the starting interval) begins, indicating the start of a new trial. Each trial runs off as described in accordance with the program, and when the last trial is finished the apparatus "homes out," plunging *S* in darkness. If *S* makes an incorrect choice, or if the trial is programmed to be nonreinforced, I_1 starts timing as soon as *S* makes his choice.

Programming is accomplished by means of five rows of toggle switches and related equipment, providing preprogramming for as many as 20 trials. Both absolute and dependent programming are possible. In absolute programming, the position of S_+ and the direction of any forced trial are programmed in advance. In dependent programming, on the other hand, either of these factors may depend on the outcome of any single previous trial. For example, one may program S_+ to be on the right on Trial 18 if *S* went left on Trial 2, or one may wish to force *S* on Trial 10 to the same (or opposite) side as his choice on Trial 7.

Spatial discrimination problems are programmed by having identical illumination occur in the two opposing arms. Finally, by always keeping one door closed, ALDA becomes an abbreviated runway.

An essential part of the programming panel is the visual display of *S*'s behavior, provided by a running-time counter and three pilot lights that simulate the maze. At the beginning of a trial the counter is reset to zero, and the light corresponding to the arm in which *S* is located is brightly lighted. As the doors open, the counter starts counting; and when *S* makes a choice, the counter stops and the light of the arm chosen goes full bright, while the light corresponding to the starting arm of that trial goes dim. Thus, *S*'s progress on every trial is easily followed from the visual display.

Multiple units of any automatic apparatus would not be feasible without some means of automatically recording the data. A Clary 11-column print-out machine is used for this purpose. The results of each trial are printed out as 11 digits, giving full information concerning the program on that trial and *S*'s choice and running time.

Each of the four Y mazes is separately housed in a sound-deadened cubicle, with white noise being fed down over each unit from a speaker in the ceiling. The laboratory, containing the vivarium and the experimental rooms, is maintained at 75° F. ± 5°, and the relative humidity is kept at 50% ± 15%. (In the present study, however, the humidifier was not yet operative and the relative humidity varied between 12% and 40%.)

Pretraining. Four days prior to discrimination learning, *Ss* were placed on a 23-hr. water deprivation schedule, food being available ad libitum. The only pretraining given in the Y maze consisted in permitting *S* to drink eight dipperfuls of water in one arm of the maze with lights out and doors closed.

Discrimination training. For all *Ss*, S_+ was the brightly lighted arm (10.0 ft.-c.), while S_- was the dimly lighted arm (0.15 ft.-c.). (These readings were taken under standard contitions, i.e., with the target placed on the grid floor, its leading edge 2.5 in. from the closed door.) The starting stimulus was a flashing light of moderate intensity. The intervals employed were: reward, 1.5 sec. ± .5 sec.; stimulus, 8 sec.; 57 sec. intertrial (lights out) ±3 sec.; and 4.5 sec. starting, ±.5 sec.

All *Ss* were run 10 trials on Day 1 and 20 day thereafter until the criterion of 18 correct out of 20 successive trials was reached. The position of S_+ was determined by a standard program, 01001001110110001101, with "0" being interpreted as left and right on alternate days.

The average time of deprivation at the start of a day's trials was 22.2 hr. with a range of 18–25 hr.

Reversal learning. Upon reaching criterion each of the 30 *Ss* was assigned to either Group R or Group ER, resulting in two groups of 15 *Ss* each matched with respect to mean number of trials to criterion. The *Ss* of Group R started on reversal learning the day after meeting criterion (i.e., the dimly lighted arm was now S_+), receiving 20 trials per day. The *Ss* of Group ER, on the other hand, were first subjected to 60 extinction trials, 20 per day, and began reversal training on the fourth day. It should be noted that the extinction trials differed from the reversal trials only in that reward was absent in the former. For convenience, these extinction trials are referred to as unrewarded reversal trials and a choice of the dimly lighted arm is called a "reversal" response. Reversal training continued until the same criterion, 18 correct responses out of 20 trials, was reached. In addition, in order to obtain a reversal learning curve, all *Ss* were run a minimum of 12 days (240 trials). The average deprivation was 22.2 hr.; range: 19–24.5 hr.

RESULTS AND DISCUSSION

Discrimination. The matching procedure resulted in mean numbers of trials to criterion of 121.5 and 123.4 for Groups R and ER, respectively. Since discrimination and reversal learning scores were significantly correlated in Group R (rho = .604, $P = .05$), the importance of some matching procedure is indicated.

Reversal. Figure 2 summarizes the reversal data. Two things are clear from the figure. First, it shows that, as compared to Group R, Group ER made considerably more "reversal" responses on Days 1–3, during which time it will be recalled Group ER was on extinction. The differences are, by the *t* test, significant for Days 2 and 3 and for the totals of Days 1 through 3 (the respective *P*s were .02, .01, and .001). The mean number of reversals for the total 60 trials was 8.8 for Group R and 14.3 for Group ER. Thus, removal of reward during the early trials of reversal learning serves to increase markedly responses to the former S_-, fulfilling the premise of the hypothesis that initiated this study.

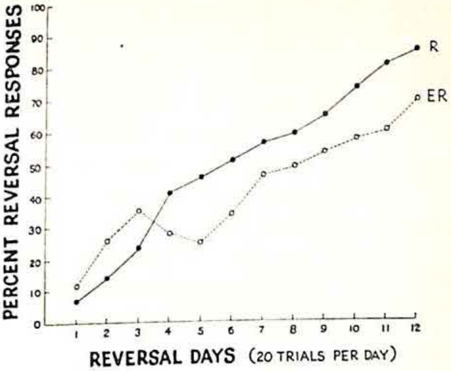


FIG. 2. Group learning curves during the first 12 days on reversal learning.

Second, however, it will be observed that as soon as reward is reinstituted, on Day 4, the performance of Group ER deteriorates, dropping below that of Group R and remaining there for the duration of reversal training. In fact, from Day 4 on the two graphs are roughly parallel with the curve of Group ER running at least two days behind that of Group R. The difference in the mean number of trials to reversal criterion reflects this displacement, being 191.0 for Group R and 243.4 for Group ER (medians: 197.0 vs. 243.0). This difference is significant by both the *t* and the *U* tests ($P < .005$ in both cases). Thus, far from facilitating discrimination reversal learning, early extinction trials severely impede such learning.

Analysis of the role of reward during reversal learning. Analysis of the role of reward during reversal learning reveals that this factor serves at least three important functions.

First, consider the incentive function of reward. In Fig. 3 the running times are graphed for the first three days of reversal learning. Each point

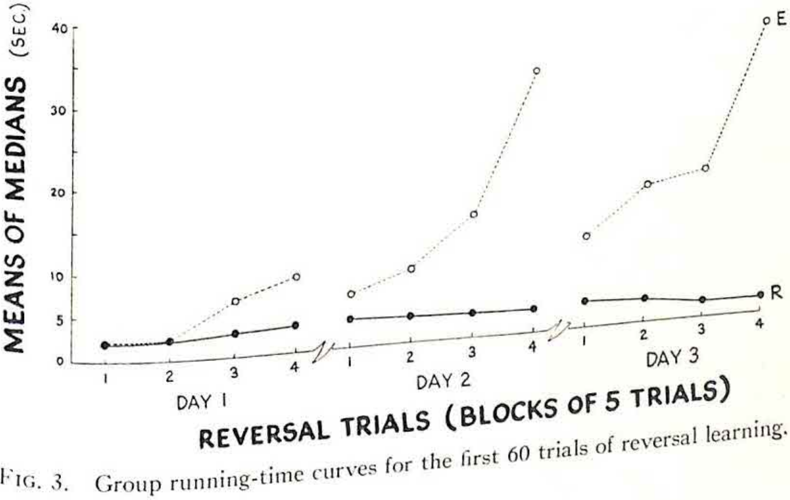


FIG. 3. Group running-time curves for the first 60 trials of reversal learning.

on the graph was obtained by finding the median running time for each block of five trials for each *S* and then taking the mean of these medians. The important features of the graph for our purposes are that: (*a*) for the first 10 trials of Day 1 both groups run equally fast; but (*b*) by the end of Day 1, Group ER slows up considerably and becomes progressively worse during the three-day period. These features show that the occasional reward received by Group R serves to keep *S*s of that group running rapidly albeit to the wrong side. The efficacy of the incentive function of reward is remarkable in view of the fact that during the first reversal day *S*s of Group R obtained only 1.33 reinforcements on the average.

Next, consider the cue function of reward. As noted earlier, on Days 2 and 3 (Fig. 2) Group R makes significantly fewer reversals than Group ER, a fact which is attributed to the presence of reward on the reversal responses of the former group.

A more striking demonstration of the cue effect of reward occurs in Group ER upon the reintroduction of reward on Day 4 of reversal learning (Fig. 2). As previously mentioned, introduction of reward on Day 4 results in a decline in the number of reversals in this group, which persists through Day 5. Statistically, the mean number of reversals made by Group ER on Day 4 (5.53) is, by the *t* test, significantly ($P = .05$) fewer than the mean number made on Day 3 (7.00), and also significantly ($P = .01$) fewer than the mean number of reversals made by Group R on Day 4 (8.27). It is concluded, therefore, that the presence of reward during the early reversal trials acts as a cue factor which serves to maintain responding that is appropriate to the initial discrimination.

The last function of reward to be considered is what may be called the *avoidance extinction function of reward*. As already seen, presence of reward during the first three reversal days leads to faster reversal learning than does absence of reward. On the other hand, the cue function of reward serves to keep *S* responding in accordance with the old discrimination. Nevertheless, something about the reversal problem is being learned, since these *S*s reverse faster than the *S*s of Group ER. The present hypothesis is that during these early reversals reward acts to extinguish *S*'s avoidance of the former S_-. The evidence for this makes use of the running-time scores of the reversal trials and of the immediately preceding trials of the first six reversals. (One *S* in Group ER made only five reversals during the first three reversal days, limiting to six the number of initial reversals available for comparative purposes.) For each of these reversals a running-time difference was obtained, i.e., the difference between the running time on that reversal and the running time of the immediately preceding trial. If reward acts to reduce the avoidance of the former S_-, then these differences should get smaller in Group R. Indeed, this is what was found. Beginning with a median difference of 4 sec. on the first reversal, by Reversal 6 the median of Group R was reduced to 1 sec. In Group ER, in contrast, the median difference was 3 sec. on Reversal 1 and reached 11 sec. by the sixth reversal. After the second reversal, the two curves do not overlap.

In summary, then, we would characterize early reversal learning as follows. By the end of discrimination training, S_+ builds up sizable approach tendencies and S_- develops rather strong avoidance response tendencies. In order to reverse, *S* must extinguish in some measure the approach tendencies to the former S_+ but even more important, he must also extinguish the avoidance tendency toward the former S_-. Reward, aside from its incentive function, serves to keep *S* approaching the old S_+, and since these responses are unreinforced, they undergo extinction. Finally, the rewarded reversal trials act to extinguish avoidance of the former S_-. In our opinion this is by far the most important function. It will be recalled that the mean difference between the groups in arriving at the reversal criterion was of the order of 52 trials. Since in Group ER the 60 unrewarded reversal training trials could only serve to extinguish the approach tendencies to the former S_+, we attribute the eight trial saving to this factor. Conversely, the 52-trial advantage of Group R over Group ER may be attributed to the few (8.8 on the average) rewarded reversals that occurred during the first 60 reversal training trials. Later in reversal learning, other factors such as partially reinforced position responding, enter to complicate the picture and protract the course of reversal learning.

We wish, finally, to comment on our application of the notions of approach and avoidance to discrimination learning. Some years ago, Nissen (1950) proposed that discrimination learning could be handled in terms of approach-avoidance response categories, a proposal that promptly drew criticism from some quarters (Weise & Bitterman, 1951). In our view this idea has great merit, though we would not necessarily defend the position that *all* discrimination learning should be so interpreted. However, application of the notions of approach and avoidance to simple discrimination learning leads to many interesting deductions and offers promise of integration of this topic with other important areas, e.g., conflict, extinction, and perhaps others. For example, we have maintained that the development of avoidance tendencies toward S_- is an important factor in the establishment of a discrimination and, conversely, reversal learning critically requires the extinction of this avoidance. Any agent, therefore, that acts to reduce avoidance tendencies should impede original discrimination learning but facilitate reversal learning. One possible class of such agents, if we are willing to generalize on the properties of avoidance tendencies, is the tranquilizers, heavily investigated in other connections.

It now seems fairly well established (Pubols, 1956; Reid, 1953) that discrimination reversal learning is *facilitated* by overlearning on the original discrimination. Our explanation is that very few errors are made during the overlearning trials, and this virtual termination of experience with S_- leads to a reduction in the avoidance tendencies (via generalization of approach tendencies from S_+, if through no other mechanism) earlier developed toward this stimulus, therefore rendering reversal learning easier. Our clear prediction is that, if *S* were forced to have a reasonable number of experiences with S_- during the overlearning trials, reversal learning would be impeded rather than facilitated. Or, equivalently, overlearning should not facilitate reversal learning in a successive discrimination situation, since avoidance of S_- is maintained by the very nature of the situation. Even these few examples may serve to illustrate the potential fruitfulness of conceiving of at least some discrimination learning in terms of the development of approach and avoidance tendencies toward S_+ and S_-.

SUMMARY

Thirsty rats were trained on a simple brightness discrimination problem in an automatic Y maze. The *Ss* of Group R were then reversed on the problem, while *Ss* of Group ER first underwent 60 extinction trials before going on reversal. Although Group ER was responding significantly more frequently to the former S_- by the end of the extinction trials (as compared to Group R at the end of its first 60 reversal trials), Group R learned the reversal considerably faster than Group ER.

Three functions of reward—incentive, cue, and avoidance extinction—were pointed out and discussed. Some implications of an approach-avoidance interpretation of discrimination learning were considered.

REFERENCES

DUFORT, R. H., GUTTMAN, N., & KIMBLE, G. A. One-trial discrimination reversal in the white rat. *J. comp. physiol. Psychol.*, 1954, **47**, 248–249.

KENDLER, H. H., & LACHMAN, R. Habit reversal as a function of schedule of reinforcement and drive strength. *J. exp. Psychol.*, 1958, **55**, 584–591.

NISSEN, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, **57**, 121–131.

PUBOLS, B. H., JR. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

WEISE, P., & BITTERMAN, M. E. Response-selection in discrimination learning. *Psychol. Rev.*, 1951, **58**, 185–195.

(Received May 11, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

LEARNING AND PERFORMANCE AS FUNCTIONS OF RATION SIZE, HOURS OF PRIVATION, AND EFFORT REQUIREMENT[1]

GLEN D. JENSEN

Northwestern University

Presumably, the psychology of animal learning will eventually reach a point of development at which exact numerical values from empirical equations fitted to parametric studies of hours of food privation (and other variables) will enter into predictive equations. The ease of making accurate predictions will vary in direct proportion to the percentage of empirical relations that are invariant over a wide range of conditions. One purpose of the present paper is to determine whether or not the relation of hours of privation to maze performance is a relation which is invariant over a certain range of conditions; namely, feeding rats various sizes of daily ration during the week preceding testing.

Ramond, Carlton, and McAllister (1955) have shown that rats which were allowed to eat a relatively small amount before privation ran significantly faster than rats which ate a relatively large amount of food before privation even though privation time was constant for both groups. Their finding implies that detailed and accurate prediction will require a *set* of empirical curves relating running time to hours of privation (one curve for each size of ration). But the findings of Ramond et al. (1955) allow only of the conclusion that the asymptotes of such curves will vary. A more interesting question is whether or not the *shapes* (e.g., slopes) of such curves will vary with ration size. If the shape remains invariant, then the concept of "hours of privation" may still turn out to be a highly "significant" one, in Bergmann's (1957) terminology (i.e., one that enters into many empirical laws in a simple manner). However, the finding that the shape changes as a function of ration size would portend the need to construct an elaborate predictive apparatus to handle motivational variables. At a minimum, it would be necessary to table pairs of slope values and asymptote values for an equation relating performance to privation time, one pair for each size of ration.

To determine whether the shape of the empirical relation between hours of privation and instrumental learning is a function of ration size, a factorial design is mandatory. A significant interaction between hours of privation and ration size would affirm that such a relationship holds. However, lack of a significant interaction would not necessarily negate the relationship.

In the present study three sizes of daily ration (6, 8, and 10 gm.) were varied factorially with three levels of food privation (1, 5, and 22 hr.). It was hypothesized that 6-gm. *S*s would show faster running speeds than the 10-gm. *S*s, with the 8-gm. *S*s showing intermediate speeds. Such a finding would confirm Ramond et al. (1955). Furthermore, it was hypothesized that ration size and privation time would interact, with the 6-gm. *S*s showing a more gradual slope than the 8-gm. and 10-gm. *S*s. That is, it was predicted that the (undernourished) 6-gm. *S*s would run almost as fast under 1 hr. of privation as under 22-hr. privation, whereas the 10-gm. *S*s would run much slower under 1-hr. than under 22-hr. privation.

---

[1] This article is based on a PhD dissertation presented to the Graduate School, Northwestern University. The author is greatly indebted to W. F. Hill, under whose guidance this research was performed. This research was supported by Grant G-2672 from the National Science Foundation.

An effort variable was included primarily to determine whether it interacts with hours of privation or ration size. If effort requirement interacts with hours of privation, then it should not be treated as a subtractive term from $_{S}E_{R}$. An attempt was also made to determine whether many closely spaced trials of a highly effortful nature would lead to depressed acquisition curves.

METHOD

Subjects and design. A total of 90 male albino rats of the Sprague-Dawley strain was assigned by table of random numbers to 45 experimental conditions. The design was a 5 × 3 × 3 factorial with 2 *S*s per cell. Three sizes of daily ration were used (6, 8, and 10 gm.), so that there were 30 *S*s in each group defined by daily ration size. There were three conditions of privation (1, 5, and 22 hr.), so that 30 *S*s underwent each condition. Finally, there were five conditions of effort requirement so that 18 *S*s were run under each of the five effort requirements. The definition of the various levels of effort requirement is treated below.

Because the *S*s came in several shipments from a stock farm, and because an analysis of variance on a previous study yielded a significant *F* between shipments, the random assignment was restricted in the following manner. Of the first 9 *S*s run, 3 were assigned to each of the three deprivation conditions and 3 were assigned to each of the three ration-size conditions. Of the first 10 *S*s run, 2 were assigned to each of the five effort-requirement conditions.

The *S*s ranged in age from 61 to 90 days with a mean of 77.9 days. In addition to the 90 *S*s which completed the experiment, three *S*s were lost for the following reasons: two *S*s failed to learn and one *S* was discarded because of a procedural error.

Apparatus. A T maze described by Cotton, Lewis, and Metzger (1958) was used with one modification: It was attached to a stand which allowed rapid changes of the angle of incline of the runway section of the maze. Two electric timers were automatically started when *S* put weight on a treadle located 2 in. from the startbox in the stem (runway) of the maze. One timer was stopped when *S* put weight on a treadle located 4 in. anterior to the choice point. The second timer was stopped when *S* touched a treadle located 3 in. from either goalbox. Arm doors prevented retracing. Time measurements were taken in .01 sec. A stopwatch was used to record the time lapse between the touch of *S*'s feet on the startbox floor and activation of the timers (the timers were activated when *S* touched the first treadle just inside the stem). This stopwatch measure will henceforth be called the "latency" measure. The values read from the two electric timers will be called "stem time" and "arm time," respectively.

Procedure. Prior to the first of seven days of "formal" pre-experimental food scheduling, *S*s were maintained in individual home cages for at least a week, during which time they were weighed at least once every two days and fed once a day. The size of the daily food ration varied with each *S* and was calculated to get each *S* as near as possible to 210 gm. of body weight by the first day of formal food scheduling. In almost all cases the *S* weighed less than 210 gm. before this initial treatment. On the first day of formal pre-experimental food scheduling, each *S* which was to be given acquisition training under the 22-hr. drive, 10-gm. maintenance condition was presented 10 gm. (±.2 gm.) of finely ground Purina laboratory chow in a plastic cup inside its home cage. Each *S* in this condition received its food 1 hr. after the time of day that it was scheduled to receive experimental treatment. From the treatment of this group, the reader can infer the treatment of the other two 22-hr. drive groups (those maintained on 8 and 6 gm. of daily ration). Regardless of ration size, *S* usually finished the ration within 20 min. Spillage was never enough to register on a scale calibrated to 0.1 gm. (This feeding procedure was chosen after extensive pilot work.)

The only difference in the treatment of the three 5-hr. drive groups from that of the corresponding three 22-hr. drive groups concerns the time of day at which *S*s were fed, and similarly for the three 1-hr. drive groups.

Each of the nine groups (10 *S*s per group) representing combinations of the three levels of daily ration and the three levels of privation time was subdivided into five groups representing five levels of effort requirement. Effort requirement was varied by changing the incline of the stem. One group was run with the maze in a horizontal position. The other groups were run with the maze tilted at angles of 2.93°, 5.80°, 8.67°, and 11.50° upwards from the startbox.

During each of the seven days of formal food scheduling prior to the first experimental day (Day 8), each *S* was handled for 5 min. a day. During the 5 min., the *S* was allowed to eat six 45-mg. food pellets from a small cup which was later to be a goalbox feeder. If *S* did not eat all of the 6 pellets during the 5 min. of taming, the pellets not eaten were placed in his home cage.

On Day 8 each of the 90 *S*s received 30 acquisition trials, each "correct" run being rewarded with a 45-mg. food pellet. The rewarded, or correct, side was assigned by table of random numbers with 45 *S*s being rewarded on each side of the maze. On each trial *S* remained approximately 10 sec. in the goalbox, after which it was placed for 15 sec. in its living cage to await the next trial. Water was available in the living cage at all times. During the first five trials, all *S*s were run with the maze in the horizontal position. The mean speed for each *S* on these first trials was later used as a predictor in obtaining adjusted means for the "effort" groups at the end of acquisition. During the remaining 25 trials on Day 8 the maze was tilted or not tilted depending upon which condition *S* was in. Usually about 10 extra sec. elapsed between Trials 5 and 6 relative to any other trials.

On Day 9 each *S* received 15 acquisition trials under the same effort requirement that obtained during Day 8. Then, after a 10-sec. delay while the maze angle was reset, each *S* received an additional 15 acquisition trials with the maze in a horizontal position. (If the effort variable had caused depressed acquisition curves for the high-effort groups, then the acquisition curves of these groups should have approached the curve of the low-effort (horizontal) group during the last 15 trials of Day 9.)

During any acquisition trial, if *S* stayed in the startbox for 2 min. without activating the timer, or if *S* did not reach a goalbox within 2 min. after activation of the timers, it was manually lifted and placed in the "correct" goalbox and allowed to eat.

Each running-time value was converted to its reciprocal to meet the assumptions of analysis of variance. All statistical testing was done on these reciprocal scores.

RESULTS

Ration size. With respect to the latency, stem time, and arm time measures, the ration-size groups appeared to approach different asymptotes. In each case the 6-gm. group approached the highest asymptotic response strength, while the 10-gm. group was always lowest, with the 8-gm. group intermediate. In the case of the stem time data and the latency data, the 10-gm. group showed superior response strength to the 8-gm. group during Trials 1–30. It is possible that the 10-gm. group was of superior initial ability on these response measures and that the ration-size variable gradually offset the difference. It was decided to test the effect of ration size by using the mean score of each *S* for Trials 31–60. If it is true that the 10-gm. group was of superior initial ability, the tests performed on Trials 31–60 should give conservative estimates of the effect of ration size. Analyses of variance were performed on the reciprocal stem time and reciprocal arm time data because the assumptions of normality of distribution and homogeneity of variance were not violated. The F between groups for the stem time data was 3.32 ($P < .05$, 2/87 *df*), and the F for the arm time data was 7.91 ($P < .01$, 2/87 *df*). A Kruskal-Wallis analysis of variance by ranks (1952) was performed on the latency data because these data were positively skewed. The H obtained was 12.07 ($P < .005$).

Since all three of the above-mentioned dependent variables show the

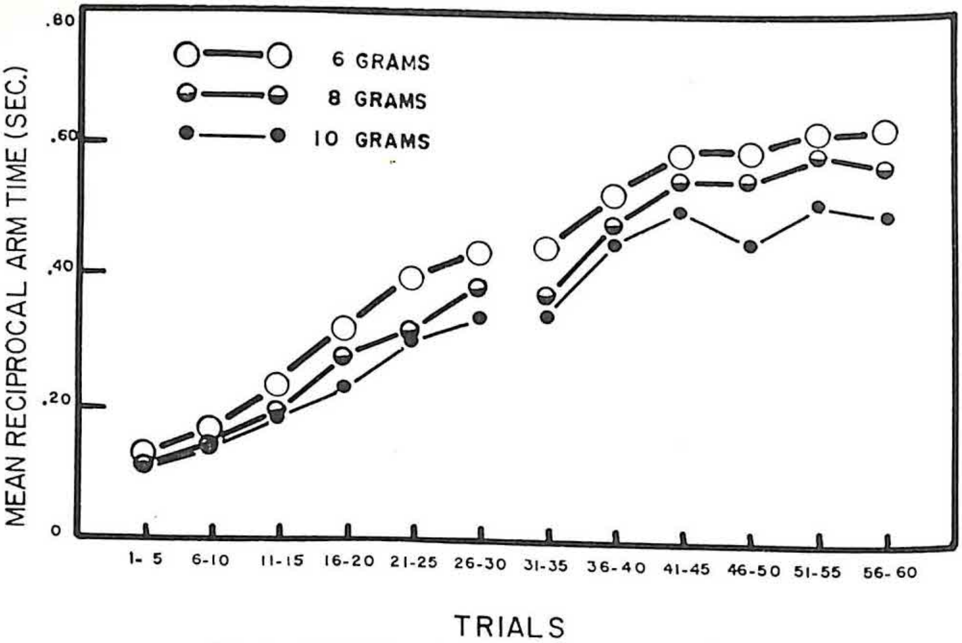


FIG. 1. Reciprocal arm time for ration-size groups.

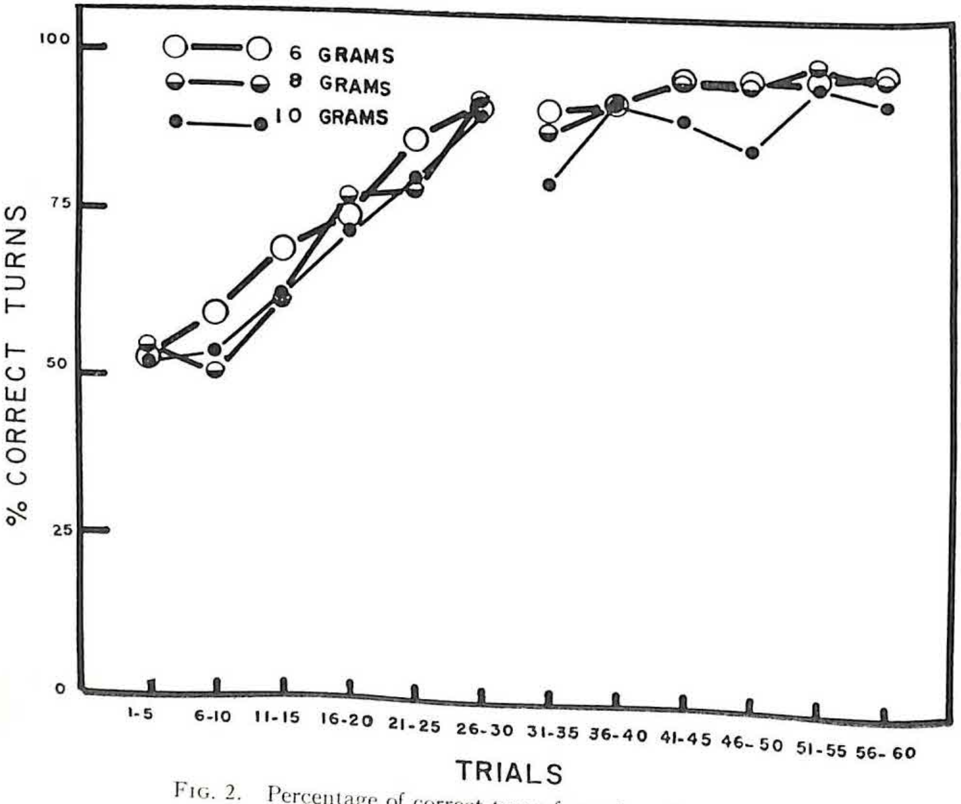


FIG. 2. Percentage of correct turns for ration-size groups.

TABLE 1

Weight and Age Data According to Ration-Size Groups

	Daily Ration Size		
	6 Gm.	8 Gm.	10 Gm.
Weight range, Day 1	207–218 gm.	208–217 gm.	205–219 gm.
Mean weight, Day 1	210.9 gm.	212.1 gm.	210.9 gm.
Weight range, Day 8	176–195 gm.	189–209 gm.	203–218 gm.
Mean weight, Day 8	183.7 gm.	197.4 gm.	211.9 gm.
Mean weight loss, Day 1–Day 8	27.2 gm.	14.7 gm.	−1.0 gm.
Age range, Day 8	61–90 days	61–90 days	61–90 days
Mean age, Day 8	77.7 days	77.3 days	78.7 days

Note.—Day 1 is the first of seven days of "formal" food scheduling. Day 8 is the first experimental day.

same general trend, only one (arm time) will be presented graphically (Fig. 1). Mean reciprocal arm time during Trials 31–60 correlated .92 with mean reciprocal stem time and −.94 with mean latency during the same trials.

The turn data (Fig. 2) do not fit the pattern of the other response measures. Note that on Trials 31–60 there is little difference between the 6-gm. and 8-gm. groups. The 10-gm. group appears, however, to be inferior to the other groups. Note that the largest disparity between the 10-gm. group and the other groups is at the data point representing Trials 46–50. Inspection of individual data revealed that most of this disparity was contributed by two *S*s in the 10-gm. group who made markedly more errors (relative to their own previous data) following the change in maze angle which occurred just before Trial 45. Therefore, in testing the effect of ration size, it appeared most conservative to exclude Trials 46–60. An analysis of variance was performed on the percentage-correct scores of each *S* for Trials 31–45. The *F* value obtained was 3.67 ($P < .05$, 2/87 *df*). Because the distribution of scores was not normal, a Kruskal-Wallis test was used to supplement the results of the *F* test. An *H* of 8.07 was obtained ($P < .02$).

The weight and age data of the *S*s broken down according to ration-size groups is shown in Table 1. Of special interest (for reasons to be presented below) is the fact that the 6-gm. *S*s lost an average of only 27.2 gm.

Hours of privation. Table 2 shows mean performance during Trials 31–60 as a function of privation time (except for the turn measure). Note that for the first three dependent variables (which are highly correlated) response strength is a direct function of privation time, whereas the percentage-correct shows no such relation. Tests exactly analogous to those used on the ration-size variable were performed on these data. Privation time had a

TABLE 2

Mean Performance during Trials 31–60 as a Function of Privation Time

Response Measure	Privation Time		
	1 Hr.	5 Hr.	22 Hr.
Reciprocal arm time	.485	.505	.567
Reciprocal stem time	.869	.916	.994
Latency	3.362	2.650	2.374
Percentage of correct turns (Trials 31–45)	91.4	90.6	91.1

TABLE 3

Mean Performance as a Function of Maze Incline

Response Measure	Angle of Incline				
	0°	2.93°	5.80°	8.67°	11.50°
Reciprocal arm time (Trials 1–45)	.322	.325	.330	.347	.300
Reciprocal stem time (Trials 1–45)	.725	.712	.705	.713	.641
Latency (Trials 31–45)	2.15	3.22	2.29	2.80	3.27
Percentage of correct turns (Trials 31–45)	91.8	88.5	90.7	91.8	92.6

significant effect ($P < .06$ or better) on each of the first three response measures, but not on the turn measure.

Effort. Table 3 shows mean performance as a function of maze incline. Note that none of the response measures showed an orderly ranking of means. Significance tests for the first two response measures gave no suggestion of a main effect of effort requirement (maze incline). The other data, being (graphically) even less suggestive of a main effect, were not analysed.

TABLE 4

Summary of Analysis of Variance of Reciprocal Arm Time Scores

Source	*df*	*MS*	*F*
Ration size (R)	2	9.528	3.24*
Hours of privation (H)	2	11.374	3.87*
Effort (E)	4	1.133	0.39
R × H	4	1.027	0.35
R × E	8	3.039	1.04
H × E	8	1.312	0.45
H × R × E	16	1.183	0.40
Between *S*s in same group	45	2.935	
Trials (T)	8	49.868	267.53**
T × R	16	0.235	1.26
T × H	16	0.247	1.33
T × E	32	0.191	1.02
T × R × H	32	0.229	1.23
T × R × E	64	0.085	0.45
T × H × E	64	0.116	0.62
T × R × H × E	128	0.157	0.84
Pooled *S*s times trials	360	0.186	
Total	809		

* $P = .05$.
** $P = .01$.

Interactions. The Kogan (1948) analysis of variance for trend, extended to handle three independent variables of classification, appeared to be the most efficient way to test for all possible interactions. Two identical tests were employed, one on the reciprocal stem time data and one on the reciprocal arm time data. In order to avoid a bias in the selection of trend points, it was decided to use all of the data obtained while all three variables were operative. This necessitated the exclusion of all trials after Trial 45. After Trial 45 the effort variable was no longer operative: so inclusion of such trials would create a bias in favor of a significant interaction of effort and trials. In each analysis nine trend points were used. Each trend point was a mean of a block of five trials. Not one of the 22 F ratios for interaction was significant or even suggestive. Sixteen of the F ratios were below 1.0 and the highest was 1.3.

Table 4 shows the results of the analysis of the arm time data. The table is presented to illustrate the technique used as well as the results obtained.

The turn data were used to test the interaction of the independent variables. The percentage-correct scores for Trials 31–45 entered into a 3×3×5 analysis of variance. Again, none of the interaction terms approached significance.

DISCUSSION

Ration size: Main effect. The evidence is clear-cut that ration size is inversely related to arm time, stem time, and latency. Thus the findings of Ramond et al. (1955) are confirmed and extended. One implication is that the asymptote of an empirical equation fitted to a parametric study of hours of privation will be a function of ration size, if any of these response measures are used.

The fact that the ration-size variable produced rather large differences in performance even though the range of ration sizes was not great (nor the resulting weight changes) indicates that it is advantageous to control this variable in animal studies in general. Control of this variable should reduce within-groups variance. Control can be greatly increased by using the feeding procedure herein employed (i.e., using a constant amount of ground food rather than a procedure which requires skill on the part of the rat to ingest a large percentage of the ration or to ingest a large quantity in a unit of time).

The present study offers some evidence that choice behavior is also a function of ration size, but the relationship is far from clear.

Interaction of ration size and hours of privation. The lack of significant interaction between ration size and hours of privation is an important and fortunate finding. The relation of hours of privation to performance has shown an invariance (in general shape of the function) over a range of conditions within which it might reasonably be expected to vary. Such invariances are highly desirable in any science because their presence makes the task of prediction much easier.

If the shape of the relation of hours of privation to performance had been shown to be a function of ration size (by a significant interaction of ration size and privation time), then the existing parametric studies of privation time would be of much less value to "future" psychologists (those attempting precise and detailed predictions of performance using exact numerical values for all relevant variables). This is true because in those studies the feeding procedure is such that it is not possible to state the amount that *S* actually ate per day.

It should be cautioned that failure to reject the null hypothesis does not confirm it. There may indeed be an interaction of ration size and hours of privation. However, if such an interaction exists, it is probably of small consequence or the interaction *F* ratio would have been above 1.0.

It should also be noted that the present study shows only that no significant interaction occurred when ration size was varied over the range of 6–10 gm. for 7 days. Presumably there is a value of ration size so low that all *S*s (regardless of privation time) would be sick or dead, so that a spurious interaction would be incurred by its use. The 10-gm. *S*s in the present study appeared to be quite healthy and gained slightly in weight, while the 6-gm. *S*s were (subjectively) quite thin by the time of experimentation. Hence, it is possible to conclude that within a reasonable range of ration size there is no interaction with privation time.

Effort: Main effect. The fact that the effort variable produced no acquisition main effect in this study places it at odds with three recent studies (Aiken, 1957; Appelzweig, 1951; Montgomery, 1951). The reason for this disparity may lie in the difference in response measures employed. Montgomery and Appelzweig may have obtained a relation between bar pressing and effort because a rat must get into a better position to make a highly effortful press than to make a light press, and time elapses while position is improved. Aiken's group differences could represent pain tolerance effects rather than fatigue or I_R effects because (presumably, at least) pushing a panel with the nose is painful.

The present study demonstrates that within a fairly wide range, effort is not an important variable in maze behavior. This is a particularly fortunate finding because there is a sense in which the effort variable is confounded in every maze experiment that employs a vigor measure. Any variable that increases the vigor of a response also increases the effortfulness of the response (the greater the vigor, the greater the kinetic energy gained, hence the greater the effort expended). The present study indicates that such confounding has probably not affected the validity of experimental claims to any serious degree.

Interaction of effort and trials. Solomon (1948) has pointed out that Hullian work-inhibition theory leads to a clear-cut prediction that if effortfulness of a response is increased while intertrial interval remains constant (and short), then depressed acquisition curves will result. This hypothesis was tested but not confirmed by the F ratio for interaction of trials and effort in two analyses. It would be interesting to see whether depressed acquisition curves can be produced by using a longer maze (constant for all groups) so that more potential energy would have to be gained on any trial for a specific angle of maze incline. It is not recommended that maze angles greater than 12° be employed because of slippage which spuriously increases running time. However, the present study does not permit of an optimistic forecast for such a research effort. Rather, the present study indicates that work-inhibition theory needs revision.

SUMMARY

Ninety rats were given 60 rewarded trials in a T maze. The design was a $5 \times 3 \times 3$ factorial with 2 *S*s per cell. Three sizes of daily food ration (6, 8, and 10 gm.) were varied over a 1-week period. There were three conditions of food privation (1, 5, and 22 hr.), and five conditions of effort requirement defined in terms of angle of maze incline (0°, 2.93°, 5.80°, 8.67°, and 11.5°).

The ration-size and hours-of-privation variables both produced significant effects upon acquisition running speed and latency of running. The ration-size variable produced a significant effect upon choice behavior, but the hours-of-privation variable did not. The effort-requirement variable gave no suggestion of a main effect. No significant interaction effects were found.

REFERENCES

AIKEN, E. G. The effort variable in the acquisition, extinction, and spontaneous recovery of an instrumental response. *J. exp. Psychol.*, 1957, **53**, 47–51.

APPELZWEIG, M. H. Response potential as a function of effort. *J. comp. physiol. Psychol.*, 1951, **44**, 225–235.

BERGMANN, G. *Philosophy of science.* Madison: Univer. Wisconsin Press, 1957.

COTTON, J. W., LEWIS, D. J., & METZGER, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

KOGAN, L. S. Analysis of variance-repeated measurement. *Psychol. Bull.*, 1948, **45**, 131–143.

KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. Statist. Ass.*, 1952, **47**, 583–621.

MONTGOMERY, K. C. An experimental investigation of reactive inhibition and conditioned inhibition. *J. exp. Psychol.*, 1951, **41**, 39–51.

RAMOND, C. K., CARLTON, P. L., & McALLISTER, W. R. Feeding method, body weight, and performance in instrumental learning. *J. comp. physiol. Psychol.*, 1955, **48**, 294–298.

SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1–40.

(Received May 13, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

STIMULUS MEANING IN STIMULUS PREDIFFERENTIATION[1]

SHEILA M. PFAFFLIN[2]

Johns Hopkins University

It has been found that pretraining with stimuli will facilitate subsequent learning of responses to those stimuli, even though the pretraining was not of such a nature that transfer would have been predicted from response generalization (Baldwin, 1954; Cantor, 1955; Foster, 1953; Gagné & Baker, 1950; Goss, 1953). Such facilitation has come to be called stimulus predifferentiation, and no explanation for it has yet been established (Arnoult, 1957; Vanderplas, 1958). The hypothesis by Miller and Dollard (1950) that labels learned as responses to stimuli subsequently act as cues which make the stimuli more distinctive, thus decreasing stimulus generalization and facilitating learning, has been advocated most often for this purpose. Unfortunately, this distinctiveness-of-cues hypothesis has difficulty in explaining the results of several experiments on stimulus predifferentiation. Highly distinctive labels did not produce more facilitation than similar labels when used as responses during pretraining (Gerjuoy, 1953). It is not even necessary to have a different label for every stimulus to obtain facilitation (Hake & Eriksen, 1955; Robinson, 1955). Yet the hypothesis would actually predict a decrement in subsequent performance due to acquired similarity when many stimuli had the same label.

Effort has therefore been directed to the study of additional variables. The one most frequently studied has been meaning, usually measured by number of verbal associations in the manner suggested by Noble (1952). The results of these studies have generally been discouraging (Dysinger, 1953; Goss & Greenfeld, 1958). Arnoult, however, found that if the label used during pretraining seems to *S*s to be appropriate for the stimulus, to "belong" to it, it may especially facilitate subsequent ability to recognize the stimulus. It would appear that perhaps it is not the meaningfulness of the labels which is crucial for stimulus predifferentiation, but rather the meaning of the label in relation to the meaning of the stimulus.

In the present experiment the stimuli (visual forms) were selected by a new criterion of meaningfulness; the consistency with which *S*s supplied the same label for the stimulus when asked to name it. The reasoning was that stimuli which had very well-defined meanings would receive the same label from most *S*s, while stimuli without well-defined meanings would not be given any single label with high frequency.

The modal label assigned by *S*s to a stimulus, regardless of its actual frequency, was considered to be the most appropriate or relevant label for that stimulus.

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements of the PhD degree at Johns Hopkins University. The author is greatly indebted to Clinton B. DeSoto, under whose supervision this study was conducted, for his guidance and helpful criticism, and to W. R. Garner for his advice in the planning of this experiment and in the preparation of this paper. The author is also indebted to James Deese for his encouragement and advice during the early phases of this experiment.

[2] Now at Bell Telephone Laboratories, Murray Hill, New Jersey.

Once the meaningfulness of the stimuli and their relevant labels had been determined in this way, a stimulus predifferentiation experiment was conducted to investigate the following three questions. Is *S*'s ability to learn responses to stimuli without pretraining related to the meaningfulness of the stimuli? How is learning influenced by the following types of pretraining: learning of relevant labels for stimuli; learning of irrelevant labels for stimuli; observation of the stimuli? Do the effects of different kinds of pretraining vary for stimuli which differ in meaningfulness?

METHOD

Selection of stimuli. Visual forms were used as stimuli since *S*s are accustomed to labeling stimuli of this sort. From an initial pool of 85 silhouette forms, 30 forms were selected for the stimulus predifferentiation experiment. The initial pool was drawn from various sources so as to provide a set that would be expected to vary widely in meaningfulness. Many of them were adapted from the nonsense forms used by Arnoult (1956b).

The forms were cut out of black construction paper and mounted on heavy white paper. They had a maximum diameter of 6 in. In the procedure used to determine consistency of labeling, each form was exposed to *S*s for 20 sec. The *S*s were instructed to write the name of each form as it appeared. The *S*s, 50 male students randomly drawn from the introductory psychology course at Johns Hopkins University, were run in groups not exceeding 10. The forms were presented in a different random order to each group of *S*s

The modal name for each form was obtained by determining the noun most frequently given in response to it. The consistency with which each form was labeled was defined as the percentage of the responses to that form which used the modal label. Three sets of forms were then selected for use in the main experiment. They consisted of those 10 forms which were labeled most consistently (average consistency = 94%); those 10 forms which were labeled with medium consistency (average consistency = 48%); and those 10 forms which were labeled with lowest consistency (average consistency = 12%). These forms are shown in Fig. 1 with their modal labels.

Pretraining with the stimuli. The kinds of pretraining used were: observation; the learning of irrelevant labels; the learning of relevant labels; and no pretraining. These four kinds of pretraining were made orthogonal to the three form classes so that 12 conditions resulted. The modal label given each form was used as its relevant label. The following 10 adjectives were used as the irrelevant labels for all three classes of forms: fresh, true, slow, better, complete, proper, weak, loud, alone, late.

The forms were photographed and made into 35-mm. slides. Labels appeared under the forms on those slides used in learning of the labels. The apparatus consisted of an Argus slide projector, a beaded screen, and a Timolite timer.

The *S*s were 120 male students randomly drawn from the introductory psychology course at Johns Hopkins University. None of them had served as *S*s in selecting the stimuli. Ten *S*s were randomly assigned to each experimental condition. The *S*s in the three control conditions received no pretraining. The remaining 90 *S*s received pretraining with the stimuli under the following procedure. They were run in groups of not more than four. Those *S*s who learned labels as responses to the forms saw the forms 12 times: 6 times with labels and 6 without, alternately. On the trials when the labels were not present, *S*s were required to write down the label for each form. Those *S*s who were to observe only were shown the forms 12 times without labels. The forms were presented for 2 sec. each under all conditions. Different random orders of presentation of the stimuli were used throughout.

Discrimination learning. All 120 *S*s then learned the discrimination task under the same conditions. The *S*s were given response boards which had two buttons on them. They were told that they were to learn which button to press in response to each of the forms which they would be shown. The forms were then presented to them without labels by means of an Argus projector. They were given 12 learning trials, and the order in which the forms were presented was randomized within each trial. The forms appeared on the screen for 4 sec. Following

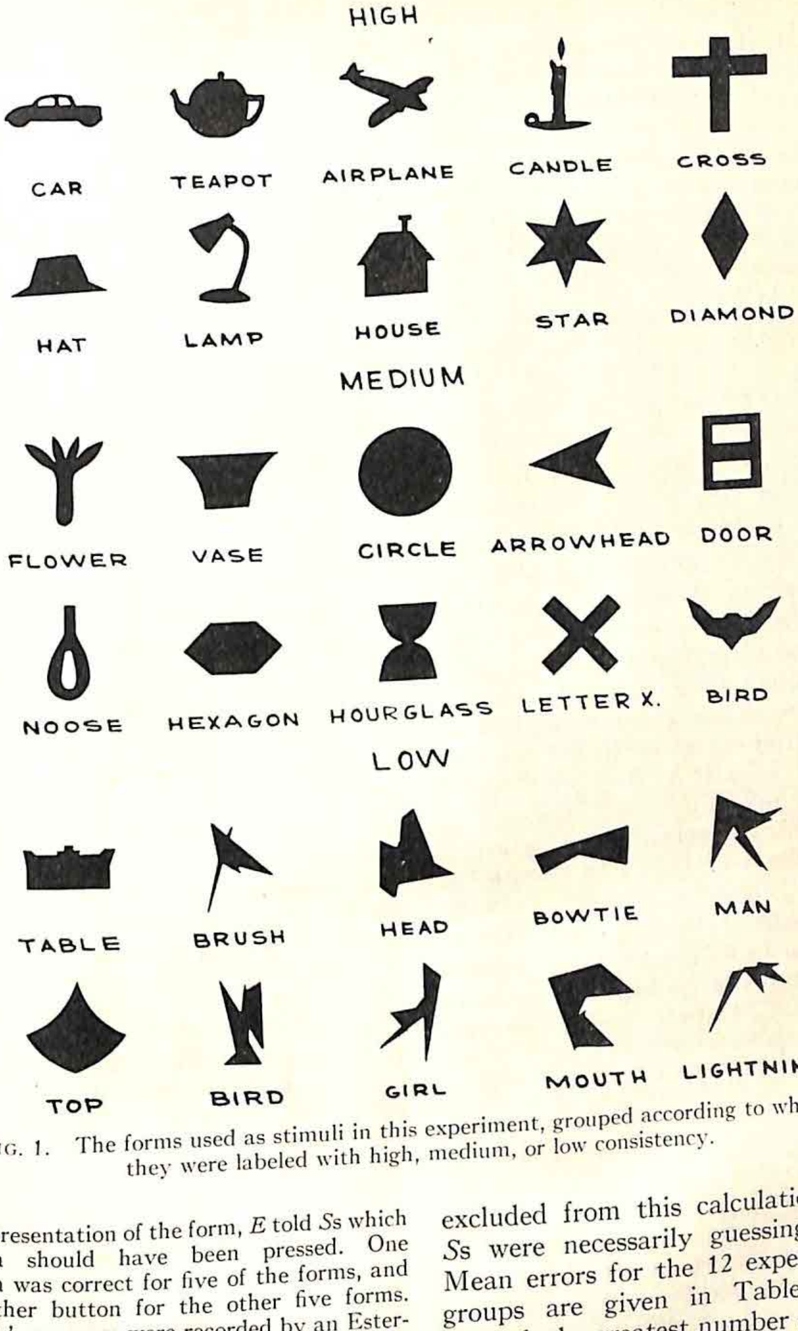


FIG. 1. The forms used as stimuli in this experiment, grouped according to whether they were labeled with high, medium, or low consistency.

each presentation of the form, *E* told *S*s which button should have been pressed. One button was correct for five of the forms, and the other button for the other five forms. The *S*s' responses were recorded by an Esterline-Angus 20-pen recorder.

RESULTS

The number of errors made by each *S* during the last 11 learning trials was determined. The first trial was excluded from this calculation since *S*s were necessarily guessing on it. Mean errors for the 12 experimental groups are given in Table 1. In general, the greatest number of errors occurred in responses to forms labeled with low consistency; the fewest in responses to forms labeled with high consistency. This relationship is reversed following pretraining by learn-

TABLE 1

Mean Number of Errors during Discrimination Learning for All Conditions

Pretraining	Form Classes			Means
	High	Medium	Low	
None	11.1	22.1	30.1	21.1
Observation	6.6	13.1	19.0	12.9
Irrelevant Labels	13.9	17.1	19.1	16.7
Relevant Labels	15.8	15.1	12.8	14.6
Means	11.8	16.8	20.2	16.3

ing of relevant labels. All kinds of pretraining result in a decrease in errors in comparison with the number made when no pretraining is given.

Analysis of variance of the error scores shows that both form classes and kinds of pretraining were significant sources of variance (Table 2). When the decreases in errors following pretraining were tested for significance by *t* test, they were found to be significant only if pretraining had been by observation or by learning of relevant labels. In view of the apparently linear relationship between consistency of labeling and number of errors, the differences between form classes were not tested for significance by *t* test. The following method was used instead.

The three form classes are approximately equally spaced on a metric, the consistency measure. The usual analysis of variance does not take this fact into account in its partition of the variance. It is possible to do so, however, by determining how much of the variance is accounted for by the linear correlation between the error scores and the form classes. The Pearson *r* was calculated by taking the intervals between form classes to be equal, and assigning the form classes the arbitrary values 1, 2, and 3. This procedure was considered to be justified since the average consistency values for the form classes depart only slightly from equal spacing, and these values were themselves subject to some chance factors in their determination. The linear component of the variance due to form classes was then calculated and found to be significant at the .01 level. There is an inverse relationship between the amount of consistency shown in labeling forms and the number of errors made in learning to respond to them. The nonlinear component is not significant, and is in fact less than would be expected from chance variation.

It is also possible to separate the interaction into linear and nonlinear components by applying the same procedure to the scores for the different kinds of pretraining separately. The linear component is significant at the .05 level, while the nonlinear component is again unexpectedly small. The entire interaction is not significant at the .05 level, though it approaches this level. Examination of the linear trend for each kind of pretraining separately shows that with

TABLE 2

Analysis of Variance of the Errors Made during Discrimination Learning

Source	*df*	*MS*	*F*
Form Classes (F)	2	738.03	6.46**
Linear Component	1	1462.03	12.88**
Nonlinear Component	1	14.04	
Predifferentiation Type (P)	3	379.46	3.32*
F × P Interaction	6	221.90	1.94
Linear Interaction	3	441.65	3.86*
Nonlinear Interaction	3	2.15	
Within Cells	108	114.31	

* $P = .05$.
** $P = .01$.

no pretraining, or with observation only, there is a significant linear component. The linear component is not significant for either of the labeling conditions.

DISCUSSION

The results of this experiment support the view that both stimulus meaning and response meaning are important factors in stimulus predifferentiation. Of particular interest is the finding that stimuli which differ in meaningfulness are affected differently by various kinds of stimulus predifferentiation. Learning responses to forms low in meaning seems to be especially facilitated by pretraining with relevant labels. Learning responses to forms high in meaning is apparently facilitated by observation as much as by any other kind of pretraining, and may actually be hindered by pretraining with relevant labels. The distinctiveness-of-cues hypothesis as now interpreted seems inadequate to explain this finding, since it provides no reason why relevant labels should be especially distinctive when learned as responses to forms low in meaning. A supplementary explanation is suggested by the "attention-to-cues" hypothesis favored by Hake and Eriksen (1955). They believe that stimulus predifferentiation results from *S*s' learning to notice the distinctive features of the stimuli. The fact that responses are learned during paired-associate pretraining is ordinarily incidental, the important aspect of such pretraining being that it forces *S*s to observe the stimuli carefully. It would follow from this view that relevant labels, if they help *S*s to select distinctive features of the stimuli, would aid subsequent learning.

The problems remain of how relevant labels help *S*s select relevant features of the stimuli, and why they do so only for stimuli low in meaning. To answer these questions it is necessary to make some assumptions about what constitutes distinctiveness. Let us suppose that people build up concepts, categories, or central processes from their past experience, and that these concepts include one or more labels. Further, let us suppose that *S*s are aided in discriminating among future stimuli by relating them to these categories. Stimuli are distinctive which are easily related to different categories. Such stimuli are also meaningful. Stimuli low in meaning are hard to relate to any category. In the absence of appropriate categories, *S*s, by observing the stimuli repeatedly, may develop new categories which will aid them in discriminating among the stimuli, but this would necessarily be slower than using already established categories. In the case of stimuli low in meaning, the relevant labels suggest appropriate categories, thus making it easier for *S*s to discriminate whatever distinctive features the stimuli have. Irrelevant labels would not have this additional selective effect, and contribute to stimulus predifferentiation only by forcing *S*s to observe the stimuli carefully. Stimuli high in meaning can be easily related to a concept without the help of relevant labels, and subsequent learning might actually be hindered by relevant labels if their repeated use induced fatigue and made the concept harder to activate. This approach seems consistent with views of the role of central processes in learning and perception that have been advocated by others (Bartlett, 1932; Hebb, 1949; Humphrey, 1951). It is also consistent with the original interpretation of the distinctiveness-of-cues hypothesis in which the term responses could include central processes associated with the stimuli. At the same time it explains why labels arbitrarily chosen as responses for stimuli without regard for the meaning of the stimuli have not produced the induced similarity or distinctiveness expected from the properties of the labels alone.

SUMMARY

Three sets of 10 visual forms each were selected so that the sets differed in meaningfulness. Meaningfulness was measured by the consistency with which 50 *S*s labeled each form. Another 120 *S*s were then

randomly assigned to one of the sets of forms. These *S*s were required either to observe the forms, to learn irrelevant labels for the forms, to learn appropriate labels for the forms, or were given no pretraining with the forms. Next these *S*s learned to press one of two buttons in response to each form in their set.

The results show an inverse linear relationship between number of errors made in learning the button pressing response and meaningfulness of the stimuli. Both observation and learning of labels appropriate to the stimuli produced significant decreases in errors below the number found without pretraining, but the linear relationship did not hold after learning of appropriate labels. The results are interpreted to support the view that the meaning of both stimuli and responses used during predifferentiation training must be considered in relation to each other before the process can be adequately understood.

REFERENCES

ARNOULT, M. D. Recognition of shapes following paired-associates pretraining. In G. Finch & F. Cameron (Eds.) *Symposium on Air Force human engineering, personnel, and training research.* Washington, D. C.: Nat. Acad. Sci.—Nat. Res. Coun., Publ. 455, 1956, 1–9. (a)

ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269–276. (b)

ARNOULT, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 339–350.

BALDWIN, R. D. Discrimination learning as a function of stimulus predifferentiation and mediated association training. Unpublished doctoral dissertation, State Univer. of Iowa, 1954.

BARTLETT, F. C. *Remembering: A study in experimental and social psychology.* Cambridge, Eng.: Cambridge Univer. Press, 1932.

CANTOR, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.*, 1955, **50**, 180–184.

DYSINGER, D. W. An investigation of stimulus predifferentiation in a choice discrimination problem. Unpublished doctoral dissertation, State Univer. of Iowa, 1951.

FOSTER, H. W. Stimulus predifferentiation in transfer of training. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

GAGNÉ, R. M., & BAKER, K. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.

GERJUOY, I. R. Discrimination learning as a function of the similarity of the stimulus names. Unpublished doctoral dissertation, State Univer. of Iowa, 1953.

GOSS, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419–427.

GOSS, A. E., & GREENFELD, N. Transfer to a motor task as influenced by conditions and degree of prior discrimination training. *J. exp. Psychol.*, 1958, **55**, 258–269.

HAKE, H. W., & ERIKSEN, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161–167.

HEBB, D. O. *Organization of behavior.* New York: Wiley, 1949.

HUMPHREY, G. *Thinking.* New York: Wiley, 1951.

MILLER, N. E., & DOLLARD, J. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

NOBLE, C. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

ROBINSON, J. S. Effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–114.

VANDERPLAS, J. M. Transfer of training and its relation to perceptual learning and recognition. *Psychol. Rev.*, 1958, **65**, 375–385.

(Received May 21, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

SUPPLEMENTARY REPORT: PROMPTING VERSUS CONFIRMATION IN PAIRED-ASSOCIATE LEARNING[1]

JOHN OLIVER COOK AND MORTON EDWARD SPITZER

North Carolina State College

Cook and Kendler (1956) and Cook (1958) have shown that Prompting (stimulus term, then response term, then overt practice of the response term) is superior to Confirmation (stimulus term, overt practice of the response term, and then the response term) in learning. The present experiment was undertaken to discover some reasons for this superiority. On the basis of a theoretical model (Cook & Kendler, 1956) plus some conditioning principles, it was hypothesized that the superiority of Prompting was due, at least in part, to a shorter between-term delay, the *omission* of overt practice of the response, or both. Since the temporal interval between the presentations of the two terms of a pair is shorter in Prompting (.25 sec.) than in Confirmation (3.25 sec.), and since the longer the S-R delay is, the poorer the learning will be, it was predicted that a short delay would yield a better learning performance than a long delay. Moreover, on the basis of the theoretical model, it was believed that the occurrence of overt practice during the delay interval would interfere with the establishment of the essential S-R connection that must take place over the span of the interval, and might also interfere with the learning of the response term as such.

Using the same materials as in the previous experiments, the hypotheses were tested by means of a 2 × 2 design consisting of the following four conditions (A) Prompting—No Overt Practice, (B) Confirmation—No Overt Practice, (C) Prompting—Overt Practice, and (D) Confirmation—Overt Practice. (The "program" in Cond. D is essentially the same as the program employed in Skinner's (1958) teaching machines.) There were 35 male undergraduates per condition. The *S*s were run to mastery with a test after every three of the 27 training trials.

TABLE 1

MEANS IN SUCCESSIVE TESTS

Tests	*C* Scores				*D* Scores			
	A	B	C	D	A	B	C	D
1	2.20	1.69	1.60	0.74	3.23	3.34	3.63	2.57
2	5.14	3.80	3.43	1.86	5.69	4.77	5.00	3.77
3	7.14	6.54	5.71	3.80	7.63	7.26	6.51	5.00
4	8.77	8.06	7.57	5.06	8.80	8.40	7.83	6.17
5	9.57	8.80	8.57	6.57	9.57	8.97	8.80	7.17
6	9.51	9.26	9.09	7.57	9.51	9.40	9.17	8.06
7	9.57	9.51	9.54	8.29	9.63	9.57	9.60	8.69
8	9.86	9.71	9.77	8.63	9.86	9.71	9.77	9.00
9	9.94	9.71	9.80	8.71	9.97	9.71	9.80	9.03

TABLE 2

SIGNIFICANCES OF DIFFERENCES AMONG MEANS (MANN-WHITNEY *U* TEST)

Tests	*C* Scores						*D* Scores					
	A > B	C > D	A > C	B > D	A + B > C + D	A + C > B + D	A > B	C > D	A > C	B > D	A + B > C + D	A + C > B + D
1	.067	.001	.007	.001	.000	.001	.295	.000	.236	.000	.134	.000
2	.010	.002	.003	.000	.000	.001	.041	.000	.119	.000	.013	.000
3	.142	.002	.020	.000	.000	.004	.166	.002	.017	.000	.000	.007
4	.115	.000	.011	.000	.000	.001	.192	.002	.014	.000	.000	.007
5	.026	.002	.017	.001	.000	.001	.029	.002	.022	.001	.000	.001
6	.359	.015	.171	.005	.005	.033	.356	.021	.179	.005	.006	.048
7	—	.055	—	.031	.078	.462	—	.048	—	.029	.085	—
8	—	.020	—	.054	.092	—	—	.020	—	.055	—	—
9	—	.192	—	—	—	—	—	—	—	—	—	—

As shown in Tables 1 and 2, the results, in terms of the number of correct items (the *C* score), bear out the predictions that Cond. A will be best (because of a short S-R delay and no intervening practice to interfere) and that Cond. D will be worst (because of a long S-R delay and because of the interfering effect of overt practice during the S-R delay interval). The relative positions of Cond. B and C were not predicted because the

---

[1] This research was supported by a grant from the Faculty Research and Professional Development Fund of North Carolina State College.

relative importance of the two variables was not known.

The two variables also proved to have a deleterious effect (Tables 1 and 2) upon the number of different legitimate responses—excluding duplicates—that a subject produces on a test (*D* score). A legitimate response is any one of the response terms used in the experiment, whether it is elicited by its proper stimulus or by some other one.

The conclusions are that: (*a*) Overt practice interferes both with learning the response term as such and with connecting it with its proper stimulus; (*b*) S-R delay interferes with the process of connecting a response to its proper stimulus, but it has no very consistently reliable effect upon learning the response term as such.

REFERENCES

COOK, J. O., & KENDLER, T. S. A theoretical model to explain some paired-associate learning data. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research.* Washington, D. C.: Nat. Acad. Sci.—Nat. Res. Coun., Publ. 455, 1956. Pp. 90–98.

COOK, J. O. Supplementary report: Processes underlying learning a single paired-associate item. *J. exp. Psychol.*, 1958, **56**, 455.

SKINNER, B. F. Teaching machines. *Science*, 1958, **128**, 969–977.

(Received May 31, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

SUPPLEMENTARY REPORT: CONTEXT EFFECTS ON ABSOLUTE JUDGMENTS OF LENGTH[1]

CARYL-ANN MILLER AND TRYGG ENGEN

Brown University

The present study is an extension of that reported by Campbell, Lewis, and Hunt (1958) on pitch. It was assumed that a further reduction in the arbitrariness of the response systems could be obtained by employing absolute judgment of line length rather than pitch.

Preliminary experiments with 99 *S*s showed that both drawing and estimating by eye yield reliable and reasonably accurate mean judgments of line lengths from 1.00 to 15.50 in. both for the very first (context free) line presented as well as for each of several lines presented in random sequences. There was only a tendency for overestimation by eye and underestimation by drawing (see also Reese et al. [1953]).

In the main experiment a systematic presentation of stimuli was used, following the procedure of Campbell et al., except that *S* made a total of 75 judgments (rather than 130), 25 in each phase and five per stimulus length in each phase. Line lengths of .75, 1.50, 3.25, 4.50, 6.75, 8.25, 10.75, 13.50, and 15.25 in. were drawn in the center of 22 × 28-in. cardboards. Extraneous cues about repetition of stimuli were eliminated by presenting a different cardboard (75 in all) for each judgment. Forty adult *S*s who were naive about the purpose of the experiment and had no special experience in estimating length were divided into two groups of 20, one (short–long group) judging the lines in the Phase 1, 2, 3 order, the other (long–short group) judging the lines in the reverse Phase 3, 2, 1 order. The lines shown in Phase 1 ranged from .75 to 6.75 in., the length being shifted gradually during Phase 2 until the range was 6.75 to 15.25 in Phase 3. The *S* was instructed to estimate by eye and report to *E* the length of each line independently and as accurately as possible to the nearest fraction of an inch.

The main results are presented in Fig. 1 where mean judgments per trial are plotted for the 6.75-in. line, the only common reference point for all three phases. Analysis of variance (Type I, Lindquist, 1953) indicates that only the Group by Trial interaction is reliable ($P < .001$, $F = 26.98$, 2 and 76 *df*). Thus context effects are indicated because judgments of the 6.75-in. line vary according to the stimulus context (or phase) within which the stimulus is presented. Differences between groups and trials are not reliable. Context effects on the other lines were in the same direction, those lengths closest to the 6.75-in. line being affected the most. As in the preliminary experiments, comparison of mean judgments and actual line lengths (.75—15.25 in.) indicated no appreciable loss in accuracy, although there

---

[1] This report is based on an Honors thesis written by the first author under the direction of the second author.

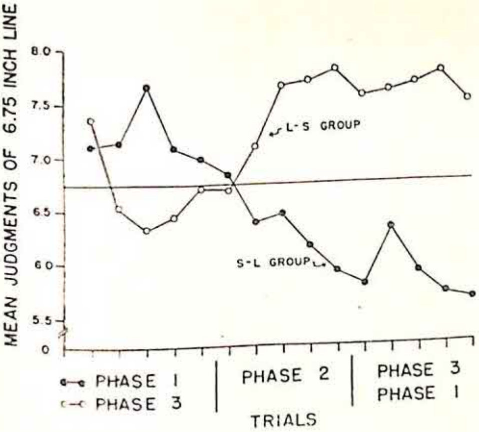


FIG. 1. Mean judgments of the 6.75 line in 15 trials distributed in three phases.

appears to be a tendency for *S*s who start with short lines to underestimate and *S*s who start with long lines to overestimate actual line length.

These results agree with those reported by Campbell et al., where comparison is feasible, in showing evidence of context unconfounded by the semantics of the response system. It should be noted that the curves obtained here cross approximately at the reference point, while in the cited pitch experiment the crossing was somewhat below the common reference point, indicating a constant error in judgment. The further minimization of arbitrariness of the response system in the present experiment appears to have improved the clarity of the results.

REFERENCES

Campbell, T. C., Lewis, N. A., & Hunt, W. A. Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *J. exp. Psychol.*, 1958, **55**, 220–228.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

Reese, E. P., Reese, T. W., Volkmann, J., & Corbin, H. H. (Eds.), Psychophysical research; summary report 1946–1952. *USN. Spec. Der. Cent. Tech. Rep. 1953, SDC-131-1-5.*

(Received June 29, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

SUPPLEMENTARY REPORT: RUNWAY PERFORMANCE AS A FUNCTION OF REINFORCEMENT SCHEDULE AND ALLEY LENGTH[1]

EDWARD L. WIKE AND ROBERT REMPLE

University of Kansas

Studies of intermittent reward upon acquisition have yielded conflicting results. A survey of recent studies suggested that 100%-*S*s ran slower than 50%-*S*s in short (13–36 in.) alleys (e.g., Goodrich, 1959), while they were equal to or faster than 50%-*S*s in long (48–84 in.) alleys (e.g., Lewis & Cotton, 1957). The present problem was to study acquisition under 50% and 100% reward in short (SH) and long (L) runways. Since alley length was varied, the results also had bearing upon within-chain delay (Besch & Reynolds, 1958).

Method. The *S*s were 48 naive Sprague-Dawley male albino rats, 90–104 days old. The apparatus consisted of a starting section, two alleys (1 ft. and 5 ft.), and a goalbox. The starting section included an 8-in. starting box, a 14-in. alley, and a hurdle, $\frac{1}{2}$ in. wide and 1 in. high, which was fastened to the floor 5 in. from the starting door. Raising the door started one timer; *S*'s breaking a light beam 12 in. from the door started another. A 13.5-in. × 8-in. goalbox joined the alley to form an L. When *S* entered the goal, *E* stopped the timers. The maze was 4-in. wide inside, with 6-in. walls; it was painted gray and covered with mesh.

On Days 1–7 *S*s were gentled and their home feeding of mash was reduced by 15-min. intervals from 1 hr. to 30 min. On Day 8, 6 to 7 *S*s explored a runway for 5 min; half were randomly assigned to Alley L and half to SH. Next, pairs of *S*s ate eight .10-gm. food pellets in the goalbox, and 15 min. later were fed for 30 min. in the home cage. On Day 9, pairs of *S*s explored the maze for 3 min. On Day 10, each *S* ate four pellets in the goalbox.

On Days 11 and 12 each *S* had one trial in the full maze. Thereafter *S* had two trials per day, spaced by 5 to 10 min. From each runway group 12 *S*s were randomly assigned

---

[1] Supported by a research grant (M-1967) from the National Institute of Mental Health, United States Public Health Service, and by the General Research Fund of the University of Kansas.

to 100% reward and 12 *S*s to 50% (24-trial pattern = RNRNNRNRRNNRRRNNNRRNNNRR). On R trials *S* had two pellets in the reward dish; on N trials the dish was empty and *S* was confined for 35 sec. Training continued for 44 trials.

Results and discussion. Starting times (from start to beam, Fig. 1) and running times (from beam to goalbox, Fig. 2) were transformed into speeds (feet per second). An analysis of variance of starting speeds revealed that Blocks of Trials and Runways × Blocks were significant ($F = 190.65$, $df = 10/440$, $P < .001$, and $F = 2.70$, $df = 10/440$, $P < .001$, respectively.) It is apparent (Fig. 2) that *S*s ran *faster* in Alley L than in SH. Runways, Blocks, and Runways × Blocks were significant ($F = 182.38$, $df = 1/44$, $P < .001$, $F = 92.29$, $df = 10/440$, $P < .001$, and $F = 12.05$, $df = 10/440$, $P < .001$, respectively). Plots of the means for first daily trials disclosed a similar pattern of results. It was expected that the 50%-*S*s would outperform the 100%-*S*s late in training in Alley SH, and be equal or inferior in Alley L. But the second-order interaction failed to be significant for either response measure, and reward schedule did not have any significant effect. Accordingly, the results for the 50%- and 100%-*S*s in Alley SH were not in agreement with those reported by Goodrich (1959).

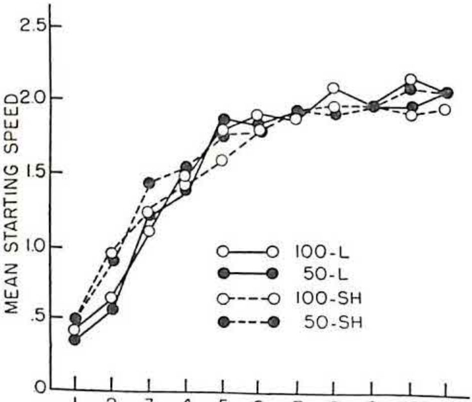


FIG. 1. Mean starting speeds.

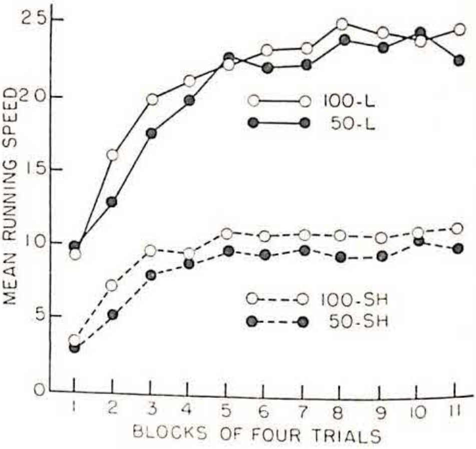


FIG. 2. Mean running speeds.

Spence's theory (1956) would predict faster starting speeds on a short runway than on a long runway. The only evidence for this prediction consisted of a significant Runway × Blocks interaction which, from an inspection of Fig. 1, is seen to be due in part to more rapid starting of *S*s in the first phase of training in Alley SH. The faster running speeds of *S*s in Alley L may reflect the fact that *S*s were unable to attain maximum speed in Alley SH. These results suggest the need for parametric studies of within-chain delay.

REFERENCES

BESCH, N. F., & REYNOLDS, W. F. Alley length and time of food deprivation in instrumental reward training. *J. exp. Psychol.*, 1958, **56**, 448–452.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

LEWIS, D. J., & COTTON, J. W. Learning and performance as a function of drive strength during acquisition and extinction. *J. exp. Psychol.*, 1957, **50**, 189–194.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received August 28, 1959)

Journal of Experimental Psychology
Vol. 59, No. 4, 1960

SUPPLEMENTARY REPORT: INFLUENCE OF INTERTRIAL INTERVAL DURING EXTINCTION ON SPONTANEOUS RECOVERY OF CONDITIONED EYELID RESPONSES[1]

ELLEN V. BEEMAN, THOMAS F. HARTMAN, AND DAVID A. GRANT
University of Wisconsin

Howat and Grant (1958) showed that with spaced extinction there were significantly fewer CRs or trials required to reach a criterion of extinction than with massed extinction. The mean spontaneous recovery was essentially the same following either massed or spaced extinction at both 20-min. and 24-hr. recovery intervals. The unexpected effect of distribution of extinction on rate of extinction made it desirable to repeat the 24-hr. recovery part of the experiment using a fixed number of acquisition and extinction trials and varying distribution during acquisition.

Method. The general apparatus, procedure, and stimuli were the same as in the earlier experiment. However, all 120 *S*s received 60 acquisition trials. For 60 *S*s trials were spaced; for the other 60 they were massed. Thirty *S*s in each of these groups immediately received 15 massed extinction trials, and the other 30 received spaced extinction. Approximately 24 hr. later all *S*s received 15 spaced extinction trials as a test of spontaneous recovery.

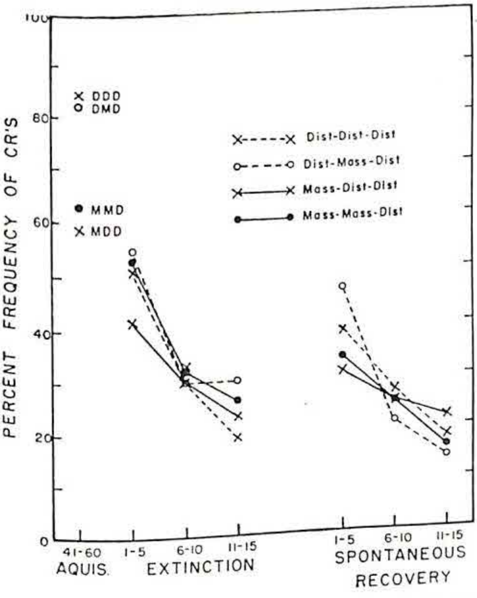


FIG. 1. Percentage frequency of CRs for all groups during the last 20 acquisition trials and on successive 5-trial blocks during extinction and spontaneous recovery test.

TABLE 1

ANALYSES OF VARIANCE

Source	*df*	*F*	
		Extinction	Spontaneous Recovery
Dist. Acq.	1	0.470	1.176
Dist. Ext.	1	1.978	0.077
Interaction	1	0.541	0.045
Error (*MS*)	116	(332.63)	(387.85)

Results. Figure 1 shows the percentage of CRs on the last 20 trials of acquisition and on 5-trial blocks during extinction and spontaneous recovery tests. The effects of distribution on acquisition were conventional and were statistically highly significant.

Table 1 gives the summary of the analysis of variance of the over-all percentage frequency of CRs during extinction, Column A, and the spontaneous recovery test series, Column B, after these percentages had been subjected to the arcsin transformation. Distribution during acquisition and extinction had no statistically significant effect upon percentage CRs during extinction or on percentage CRs during the spontaneous recovery test trials. Although there is some indication in Fig. 1 of greater spontaneous recovery for the groups receiving distributed acquisition, this result was not statistically significant.

In view of the very large number of cases and the high power of the resulting *F* tests, it is safe to conclude that distribution during extinction has little effect upon initial extinction nor upon spontaneous recovery. In the first respect the results of the earlier experiment are not confirmed. The present experiment agrees entirely with the first experiment in finding no important effect of the distribution of original extinction upon 24-hr. spontaneous recovery. Although ini-

---

[1] This research was supported in part by the National Science Foundation and in part by the Research Committee of the Graduate School of the University of Wisconsin with funds provided by the Wisconsin Alumni Research Foundation.

tial spontaneous recovery trials favored the massed extinction groups as before, the results for the massed and distributed groups were essentially identical on the total spontaneous recovery test, and the very sensitive F test was not significant.

It is not clear precisely what is the source of the difference in results of the two experiments, but it might be noted that in spite of a smaller number of *S*s in the earlier experiment, using trials to a criterion and CRs to a criterion during acquisition, extinction, and spontaneous recovery testing resulted in smaller error variances so that the earlier experiment was actually more sensitive than the present study.

REFERENCES

Howat, M. G., & Grant, D. A. Influence of intertrial interval during extinction on spontaneous recovery of conditioned eyelid responses. *J. exp. Psychol.*, 1958, 56, 11–15.

(Received August 7, 1959)

Journal of
Experimental Psychology

Vol. 59, No. 5 May 1960

THE GENERALIZATION OF ATTITUDE CHANGE WITHIN A SERIAL STRUCTURE

HELEN PEAK, H. WILLIAM MORRISON, AND R. P. QUINN[1]

University of Michigan

This study is one of a series devoted to exploring the generalization of change in attitude or affective responses within different psychological structures. It has often been assumed that antecedent–consequent patterns, such as are characteristic of a serial structure, are involved in so-called secondary or mediated generalization, and one of the unsettled issues about mediated generalization is the question of the relative importance of factors operating at the time of response training and at the time of testing for generalization (Razran, 1949).

If every act depends on the combined effects of a specific input pattern (from external and internal sources) activating an existing structure, then the response made to any stimulus being tested for generalization will be influenced by these factors. Consequently, anything which at the time of response training alters the psychological structure of which the stimulus-to-be-tested is a part will influence the appearance of generalization. Likewise any alteration in the total input to that structure at the time of testing may alter the pr ability of the occurrence of a part generalized response.

Thus if a stimulus which is previously learned serial st presented and paired with it is likely to elicit wit quency each other "sti has been learned as pa series. As a consequ sponse following the sti ing will become associ other stimuli as well. in associative structu occurred as a result of ing and generalization in the manner to Moreover, when a test s is part of such a seria presented for rating, the is most likely to be acti forward associate of the te

---

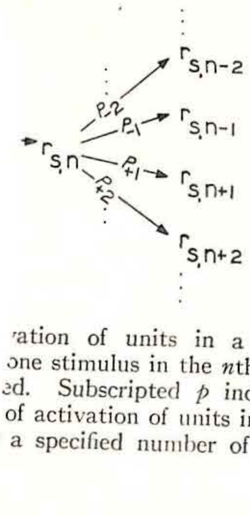
[1] This investigation was carried out under a Contract between the Office of Naval Research and the University of Michigan, Project Nonr 1224(10)NR171-039. The senior author was in residence at the Center for Advanced Study in the Behavioral Sciences and the second author was a Public Health Service Research Fellow of the National Institute of Mental Health when the article was written. Margaret Clay and Vera Schlesinger helped in the collection of part of the data.

It will be suggested that this and other events interact to determine the result at the time of testing for generalization.

In the experiment reported here a serial structure was established by rote learning of 12 nonsense syllables during relations training. After completion of this phase of training the seventh stimulus in the series (S_7) was associated with shock during response training. At the beginning and end of the experiment *S*s rated syllables S_4 through S_{10} on scales so that change in the judged affect of the trained stimulus (S_7) and six test stimuli could be measured. In a Paired-Associates Control (PA) condition, instead of learning the serial list, *S*s learned the same syllables as responses to numbers; other procedures were the same as in the serial condition.

THEORY AND HYPOTHESES

The serial structure.—In the present experiment generalization of response based on ocesses occurring either during response ing or during tests for generalization ls on the nature of the serial structure. ll known that this structure is not a quence of related units but rather a f backward and forward associa- lifferent degrees of remoteness



ation of units in a serial one stimulus in the *n*th posi- ed. Subscripted *p* indicates of activation of units in each t a specified number of steps

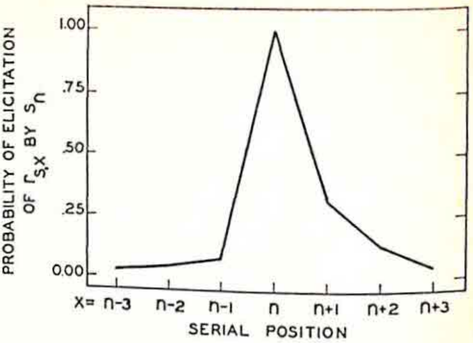


FIG. 2. Probability of activation of units in a serial structure when the *n*th stimulus is presented (estimated from Raskin and Cook, 1937).

Whenever some stimulus (S_n) within the series is presented, it is assumed that an implicit event ($r_{s,n}$) is always elicited. It is further assumed that following the elicitation of $r_{s,n}$ there is some probability of the elicitation of each other unit in the series ($r_{s,1}$ to $r_{s,12}$). These events are represented in Fig. 1, which omits activation of r_s's from sources other than $r_{s,n}$. The probabilities of the activation of these events forward and backward from some one point in a series may be approximated from the Raskin and Cook study (1937), which reports the frequencies with which serially associated nonsense syllables in various list positions were given as associates to syllables in other positions. Figure 2 represents the relative frequency of response observed at various list positions relative to the position of the stimulus syllable.[2] In the PA Control

---

[2] This study has been criticized because of methodological imperfections (Bugelski, 1950). It should be noted, however, that for our purposes the fact that the same nonsense syllable was always used in the same serial position does not raise the problems to which Bugelski refers. Every stimulus but the last was the point of origin for a one-step forward associate and every stimulus but the first was the origin for a one-step backward associate, so the frequencies at a given position relative to the stimulus syllable were not dependent on responses to one stimulus. Moreover, their technique of asking for free associations to each syllable as a means of determining its probability of occurrence was a more appropriate method for ascertaining the relative frequencies of response than the Bugelski method of counting intrusions in the course of learning by the anticipation method.

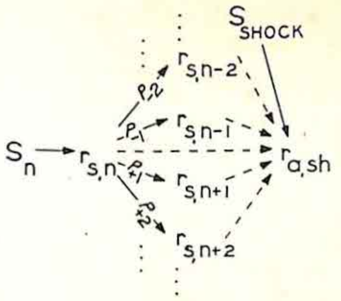


FIG. 3. The association of serial units with the affective response to shock ($r_{a,sh}$) when S_7 (S_n) is presented. A solid line indicates a previously learned association. A broken line indicates the formation of a new association at the time of response training. The contiguous occurrence of any r_s and $r_{a,sh}$ results in learning an affective response to that r_s.

condition, on the other hand, it is assumed that such relations between nonsense syllables are weak and equal in strength.

Effects of response training.—From the results described above, it may be inferred that when S_7 is presented and followed by shock, each other unit of the serial structure has some probability of elicitation and of being associated with shock. This probability depends on the list position of a given unit relative to S_7. Although never presented during response training, these implicit "stimulus" events as well as the trained stimulus will be contiguous with the negative affective reaction ($r_{a,sh}$) elicited by the shock. The frequency of association of each unit with $r_{a,sh}$ is one determinant of the affective judgment of that stimulus when later tested for generalization. Figure 3 represents the association of serial units with shock during training and the strength of negative affective response to each unit at the end of response training.

We have made a straightforward use of the Raskin and Cook data in estimating the relative strength of affective responses to each serial unit, assuming that relative response strength is a bilinear function of (*a*) relative frequency of elicitation of a given unit by S_7 and its consequent association with shock, and (*b*) the amount of change in the affective response to the trained stimulus as the result of pairing with shock. The results of such estimates are presented in Fig. 4. Three levels of trained change are illustrated. In a maximum change group, change in S_7 is given the arbitrary value of -1.00; also illustrated is a moderate change group with S_7 change represented as 50% of S_7 maximum (i.e., -0.5) and a minimum change group with S_7 equal to 25% of S_7 maximum (i.e., -0.25). Obviously, for different *S*s change in S_7 may be any proportion of the maximum change value for S_7. In order to estimate the relative affective reaction to stimuli in each serial position in groups with S_7 change 100, 50, and 25% of S_7 maximum change, we have multiplied such values of S_7 by the probability of elicitation of any $r_{s,n}$ by S_7 taken from Fig. 2. Because of the shape of the curve in Fig. 2, the affective reaction to S_8, the forward associate of the trained stimulus, is more negative than that to any other untrained stimulus and there is a gradient falling off in both directions from S_7.[3]

On the basis of this analysis Hypotheses 1 and 2 were proposed:

1. The amount of negative change in the rating of S_8 decreases with decrease in the amount of negative change in S_7.

2. When the rating of S_7 changes most negatively, the rating of S_8 changes (*a*) more negatively than ratings of untrained stimuli in the PA Control condition, and (*b*) more negatively than ratings of syllables at all other serial positions except S_7.

Consideration of processes occurring at the time of testing do not affect the above predictions about generalization of response to S_8 (see below).

Effects at the time of test for generalization.—After one-third of the cases had been run it became apparent that our predictions were being fulfilled in that change in the rating of S_8 was more negative than that of other syllables in the list and more negative than PA Controls. However, it was also clear that the change in rating of S_6 did not conform to the prediction made on the basis of events during response training; namely that change in the rating of S_6 would be negative though not as negative as S_8 (see Fig. 4). Nor was the result the one which should have been predicted by Razran (1949) or by other proponents of a mediation theory (Osgood, 1953); presumably they would predict that the stimulus antecedent to a trained stimulus in a serial list would produce the most negative generalized response. Compared with appropriate controls the rating of S_6 changed positively in the group which showed maximum negative change in S_7. There was a less striking tendency for

---

[3] There is a question about the probabilities of arousal of syllables at the ends of the serial list but these are not involved here.

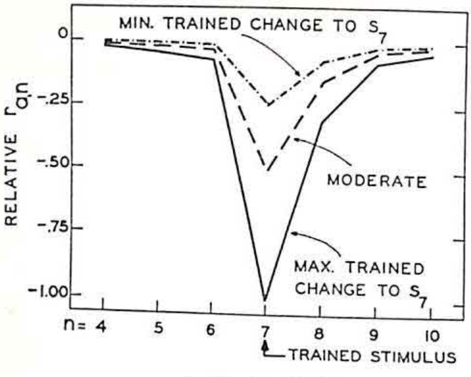


FIG. 4. Relative strength of implicit affective response to stimuli in different serial positions at three levels of change in response to the trained stimulus.

the rating of S_6 to change negatively with moderate negative change in the rating of S_7.

About the time that these results were obtained, Mink (1957) reported an experiment and a theory suggesting that the appearance of a generalized response is determined in part by the interaction between events occurring at the time of testing. Using paired word stimuli with much stronger associative strength from the first to a second word than in the reverse direction, he observed generalization of a voluntary finger response from the first to the second word, and compared this with generalization when the response was associated with the second word. The former arrangement yielded significantly more generalization. In the latter arrangement Mink attributed the absence of generalization in the direction of the trained response to *S*'s opportunity to compare the (first) test stimulus with the implicitly elicited trained "stimulus." Mink suggested that when *S* compared the tested and trained words he would realize that they were different and therefore would not respond to the test stimulus even though there was some tendency to do so.

One of us had previously elaborated a theory of activation within various kinds of structures which developed the notion that the occurrence of assimilation and contrast is a function of the psychological distance between two events,[4] and it became apparent at this point that the behavior of S_6 might be accounted for by contrast and assimilation effects between the complex of events elicited at the time of testing for generalization. The theory explained below shares with Mink's the notion that generalization is a function of events occurring at the time of testing and at the time of response training. It goes beyond Mink's formulation chiefly in stating more explicitly some of the conditions and consequences of such interactions between elicited events.

It is assumed that in our experiment the evaluative rating given to any syllable will be the result of the interaction of the various tendencies to make affective responses of some sign and intensity to the test syllable and to the other events activated on such an occasion. Again the properties of a serial structure are critical, for when a stimulus is presented for rating, the probability of elicitation of other serial units (and hence of their affective response tendencies) is determined as in Fig. 2. It should be noted that when S_n is presented, $r_{s,n+1}$, its forward associate, has a much higher probability of elicitation than other units in the list. In deriving predictions from processes at time of testing for generalization we do not consider elicitation of list units other than $r_{s,n+1}$ and $r_{s,n}$ by S_n. Figure 5 illustrates this simplified schema of the situation at the time of testing.

The test syllable, S_n, elicits $r_{s,n}$ and its associated effective response ($r_{a,n}$). At the same time the next serial list unit ($r_{s,n+1}$) has a high probability of being elicited, and will itself produce an affective response ($r_{a,n+1}$). The affective responses associated with each serial unit are those resulting from response training (see Fig. 4). The two levels of affective response ($r_{a,n}$) and ($r_{a,n+1}$) together with the input from the instructions to judge S_n, will interact to determine the overt affective judgment.

As already suggested, this judgment is assumed to be a function of the difference or distance between $r_{a,n}$ and $r_{a,n+1}$. Where there is a small absolute difference between

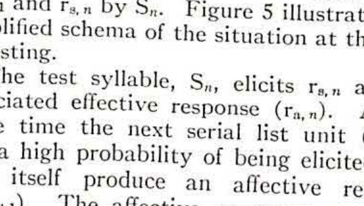


FIG. 5. Events at Time 2 rating of stimulus (S_n).

---

[4] This theory was discussed at the ONR Symposium at Harvard in March 1957. A revision of the paper presented at that time appeared in Tagiuri and Petrullo (Peak, 1958a). The theory is developed further in Peak (1958b).

these values, assimilation will occur and the judged affect of S_n is closer to that value corresponding to $r_{a,n+1}$ than if $r_{a,n+1}$ had not been elicited. At greater distances, contrast will result. In this case, the judged affect of S_n will be reduced when $r_{a,n} < r_{a,n+1}$ and increased if $r_{a,n} > r_{a,n+1}$, compared with the situation where $r_{a,n+1}$ is not elicited. Since the derivation of difference in affective responses for different list positions is relative, and since the absolute distances between events which result in assimilation or contrast are not known, predictions about these effects must be made in relative terms.

Following this paradigm, S_6 as a test stimulus will elicit not only its own affective response ($r_{a,6}$) but also its forward associate ($r_{s,7}$) which in turn produces $r_{a,7}$. Since the difference between $r_{a,6}$ and $r_{a,7}$ is greater than that for any other adjacent list positions (see Fig. 4), contrast is most likely to occur when S_6 is the test stimulus being rated. Furthermore, the more negative is $r_{a,7}$, the more likely it is that this difference will become great enough to produce contrast, since the difference in level of $r_{a,7}$'s between trained change groups is greater than the corresponding difference in $r_{a,6}$'s (see Fig. 4). The effect of such contrast between $r_{a,6}$ and $r_{a,7}$ will be to make the judgment of S_6 more positive.

Hypothesis 3 follows from these considerations: When negative change in affective rating of S_7 is greatest, the affective rating of S_6 will change (*a*) more positively than untrained syllables in the PA Control condition, (*b*) more positively than other syllables in the serial list, and (*c*) at the most negative level of trained change the rating of S_6 will change more positively than the rating of S_6 when there is less change in S_7.

Further consideration of the relations between $r_{a,6}$ and $r_{a,7}$ as $r_{a,7}$ becomes less negative suggests that when this difference decreases below the contrast level, negative change in the rating of S_6 will result, producing a more negative change in S_6 than in any other syllable except S_7 and S_8. This follows from the fact that S_4, S_5 and any other syllables antecedent to S_6 will activate r_a's of smaller magnitude than does S_6 (see Fig. 4), so that even if one or the other of these r_a's is assimilated to its forward associate (viz., $r_{a,5}$ or $r_{a,6}$) the judgment of S_4 or of S_5 would not be as negative as would the judgment of S_6 which would assimilate to the more negative $r_{a,7}$. Likewise, syllables following S_8 will produce smaller r_a's than does S_6, for if assimilation occurs to the forward associate of either S_9 or of S_{10} this will make the judgment of each still less negative. This leads to *Hypothesis 4:* As change in negative affective rating of the trained stimulus decreases, change in the rating of S_6 will be (*a*) more negative than change in ratings of untrained syllables in the PA Control condition, and (*b*) more negative than change in any other stimuli in the serial list except S_7 and S_8.

Considering the effects of events at the time of testing on Hypotheses 1 and 2, it is clear that as $r_{a,7}$ becomes more negative, the distance between $r_{a,8}$ and $r_{a,9}$, its forward associate, increases. This would lead to increased contrast or decreased assimilation depending on the distance. Since this would have the effect of making the judgment of S_8 still more negative, Hypotheses 1 and 2, which are concerned with the greater negative change in S_8 when S_7 change is maximum, do not need modification.

Finally, the possible influence of these interaction processes on the judgment of S_7 should be mentioned. Figure 4 indicates that as $r_{a,7}$ becomes more negative the difference between $r_{a,7}$ and $r_{a,8}$ will be greater. This will increase the contrast between them, or decrease the assimilation which would make judgment of S_7 even more negative. It is clear, therefore, that judgments of S_7 would continue to reflect the greater negativity of $r_{a,7}$ as compared with other r_a's in spite of interaction of events at the time of rating.

METHOD

Subjects

Random assignment of 175 undergraduate women was made to the following conditions: Serial (72 *S*s), Paired-Associates Control (73), and Test–Retest Control (30).

Design and Procedure

Table 1 presents the major features of the design and procedure. The *S*s in each of the three different experimental conditions performed the tasks represented in the table and described below.

Affective ratings: Time 1.—All *S*s rated 11 stimuli (2 meaningful words and 9 nonsense syllables chosen to minimize the number of common letters) on 12 semantic differential scales (Osgood, Suci, & Tannenbaum, 1957). The stimuli rated were FOG, STEEL, FAM, VED, CEF (always rated in Positions 1, 2, 3, 5, and 11, respectively, and ZUT, WIX, ROJ, YOF, BIP, and KAQ, whose rating positions varied within each group of six *S*s so that each appeared equally often in each rating position. In the Test–Retest Control the rating positions of all stimuli except FOG, STEEL, FAM, and CEF were randomized.

TABLE 1

SUMMARY OF DESIGN AND PROCEDURE

Experimental Groups	Time 1 Rating	Relations Training	Response Training	Time 2 Rating	Analysis Groups: Amount of Change in S_7
Serial	Rate 11 stimuli	Serial learning 12 syllables	VED (S_7) associated with shock	Rate 11 stimuli	Maximum change Moderate change Minimum change
Paired-Associates Control	Rate 11 stimuli	Each of 12 syllables learned as responses to numbers	VED associated with shock	Rate 11 stimuli	Maximum change Moderate change Minimum change
Test–Retest	Rate 11 stimuli	No training Interpolated tasks		Rate 11 stimuli	Used to determine regression and to correct change scores

The descriptive terms which labeled the semantic differential scales were SMALL–LARGE, BAD–GOOD*, CALM–AGITATED, THICK–THIN, NICE–AWFUL*, FAST–SLOW, DIRTY–CLEAN*, STRONG–WEAK, PLEASANT–UNPLEASANT*, DULL–SHARP, SAD–HAPPY*, AND HOT–COLD. The scales were in the above order on the page. Each continuous line extended the width of a 9-in. sheet of paper and the only subdivisions marked the positions of the descriptive terms (e.g., *Very Good* and *Very Bad*) which were placed 1.5 in. from each edge. Scores were determined by measuring the distance between the unfavorable end of the evaluative scales and the point checked to the nearest $\frac{1}{8}$ in. Scores were summed over the five evaluative scales marked with an asterisk.

The *S*s were introduced to the rating task as a study of the "variety of meanings and connotations which stimuli have for different people," and were urged to work rapidly. Each stimulus was pronounced by *S* as it appeared in the aperture of a manually operated exposure apparatus. When ratings were completed, *S*s were told that they would be asked to rate another group of stimuli later.

Relations training: Serial condition.—This task was introduced as a test of speed and accuracy of learning and *S*s were instructed in the use of the anticipation method. The *S*s learned a practice list of six nonsense syllables to a criterion of three successive correct trials and then an experimental list of 12 nonsense syllables to the same criterion. A unit was presented on a memory drum every 1.5 sec. with an intertrial interval of approximately 6 sec. The serial position of untrained stimuli-to-be-tested for generalized negative affect was varied systematically among Positions 4 through 10. VED was always in Position 7. When the task was completed *S*s were told they would return to the list later to see how well they could remember what they had learned.

Relations training: PA Control.—The *S*s in this group learned a practice list of six number–nonsense syllable pairs to three successive correct trials and then a list of 12 number–syllable pairs which were: 23-WIX, 29-BIP, 37-YOF, 43-DUZ, 51-NOL, 59-ZUT, 67-VED, 71-DEH, 79-PAF, 83-GAN, and 91-KAQ. Each number appeared for 2 sec., followed by the paired number and syllable for 1 sec. The syllables were the same as those in the serial lists and each pair appeared an equal number of times in random order.

The number of trials given to the pairs in the main experiment for each group of six PA Control *S*s was equal to the median number of trials to criterion for the preceding group of six serial *S*s. This criterion was used in order to insure that the test syllables would be equally familiar to Serial and PA Controls as a result of approximately the same number of exposures to the syllables. These Serial and PA Control groups of six *S*s were also matched on the order in which the syllables were rated at Time 1 and Time 2.

Test–Retest Control.—The *S*s in this group received neither relations nor response training. After completing the first rating of syllables, another *E* administered the following tasks which required about an hour and a half, approximately the same amount of time devoted to relations and response training in the other groups: two parts of the Cooperative School and College Ability Tests of the Education Testing Service; the 100 stimulus words of the Kent-Rosanoff list to which *S*s responded with free associations; a task requiring judgments of difference or similarity between pairs of meaningful words and pairs of shapes; and a test involving the production of synonyms and antonyms of a list of words. After the tasks were completed, Ss made a second rating of the words and nonsense syllables (see below).

Response training.—A single syllable (always VED) which had been included in relations training, and other nonlist stimuli were associated with shock in both the Serial and PA Control conditions

Two telegraph keys were placed in front of S who was told to hold down the left key until a "word" was presented in the frame. When the "word" appeared she could either (*a*) continue to press the left key or (*b*) release the key and press the right-hand key (using one hand). She was told that she would be paid for each correct response and that she might accumulate between $1.50 and $2.00 which she could keep.

The *E* then explained that two kinds of stimuli would be presented. For *punishment* stimuli *S* received a shock regardless of what she did, but by pressing the right-hand key she could turn off the shock. However, in order to be paid on that trial she had to release the left key *before* the shock was turned on, approximately 1 sec. after the presentation of the "word"; the shock continued until *S* pressed the right-hand key. For *nonpunishment* stimuli *S* never received a shock but was paid only if she did *not* release the left key.

Before training began the intensity of shock was adjusted to a level reported to be "painful." The *S* was told that at any time she could request a reduction in the intensity of shock, but that the purposes of the experiment required that it be "a punishment." During training, shock intensity was increased until *E* judged from *S*'s reactions that the shock was painful. At or before this point Ss often requested a decrease in shock, from which decreased level the program of gradual increases was resumed. Recognition of change in intensity was explained to *S* as resulting from "increased sensitivity." If she refused to be shocked or became upset the experiment was terminated. The shock was administered through a cuff fastened to *S*'s nonpreferred arm and was generated by a Phipps-Bird inductorium or by a Harvard Apparatus Company inductorium, No. 9358.

All Ss included in the results received at least 84 response training trials, 42 of which were punishment trials. The order of syllables was random within each set of five or six trials. The syllable VED, always associated with shock, and CEF, never associated with shock, both occurred once within each set of trials. Fourteen other syllables were presented, some followed by shock and some not. In addition to VED the following syllables were associated with shock: MEC, GUH, GIH, MUH, HES, MIG, NAH. In addition to CEF, the following were not related to shock during response training: LIH, HUC, HEG, NIH, CUH, NEH, GAH. If *S* responded correctly to both VED and CEF on five successive sets of trials, training was stopped. If *S* had not met this criterion within the first 84 trials the last four sets of trials were repeated until either (*a*) the above criterion was met; (*b*) VED and CEF were responded to correctly on four successive sets beginning with Trial 60, 85, or 109; or (*c*) 133 trials had been completed, whichever occurred first.

Affective rating: Time 2.—In all conditions the same nonsense syllables which had been rated at Time 1 were presented in the same order, preceded by two different meaningful words. The Ss were told before the ratings that some words might be the same as those rated before but that most would be different. In Serial and PA Control conditions the shock cuff was left on *S*'s arm during these ratings and the inductorium continued to hum, but Time 1 and Time 2 rating conditions were the same in other respects.

Reliability of ratings was determined by data from the test–retest group. When each *S*'s rating of each of the seven nonsense syllables was treated as an independent case, the correlation between Time 1 and Time 2 ratings was .448. Since for 30 Ss there was a correlation of .516 between mean Time 1 rating of seven syllables and mean Time 2 rating, this estimate of individual syllable rating reliability was based to some extent on between-individual rather than between-stimulus variance. Therefore, in order to estimate individual syllable reliability, scores were normalized for each *S* within each time of rating, and the resulting correlation between Time 1 and Time 2 *z* scores was .407. These reliabilities are lower than those usually reported for semantic differential ratings of meaningful stimuli. Both the use of nonsense material and of a continuous scale probably contributed to the lower reliability. Reliability was intentionally sacrificed since it was expected that memory of the ratings would be reduced with an unmarked scale. It was hoped that these procedures would result in increased sensitivity of the measure.

Definition of trained change groups.—Since it was predicted that generalization would be a function of amount of change in S_7, Serial and PA Control Ss were sorted in the following manner: (*a*) those who did not change their rating of the trained stimulus or who changed it positively were not used in the major analyses; (*b*) three groups of approximately equal size and differing in the amount of corrected negative trained change constituted the sample used unless otherwise indicated. These were designated as the *maximum trained change*, *moderate trained change*, and the *minimum trained change* groups (see Table 1).

RESULTS

Before proceeding to the tests of hypotheses two questions will be considered:

1. How well were interfering variables equated?

2. Was the necessary condition for a test of generalization met; that is, did the rating of the trained stimulus become more negative following its association with shock? After these questions are answered, the hypotheses will be tested by three types of analysis: (*a*) a comparison of change in affective ratings of stimuli three positions backward and forward from the trained stimulus in the Serial condition with mean change in ratings of untrained stimuli in the PA Control condition; (*b*) a comparison of change in ratings of test stimuli in each serial position with all other positions within the Serial condition; and (*c*) a comparison of change in ratings of each test stimulus in a given serial position with a stimulus in a corresponding position at other levels of trained change.

Equating Variables

In order to increase the confidence with which results can be attributed to the experimental variables (experimental condition, serial position, and amount of change in the trained

TABLE 2

EQUATION OF VARIABLES

Variable	Difference between Serial and PA Conditions at Each Serial Position within Trained Change Groups			Difference between Trained Change Groups at Corresponding Serial Positions			Difference between Serial Positions within Serial Condition and Trained Change Group		
	Max.	Mod.	Min.	Max.-Mod.	Max.-Min.	Mod.-Min.	Max.	Mod.	Min.
Mean Time 1 rating of untrained stimuli at each serial position	ns	ns	ns	ns	ns	ns	S_4-S_7 >0 $P<.05$	S_8-S_7 >0 $P<.05$	ns
Mean Time 1 rating of trained stimulus	ns	ns	ns	>0 $P<.05$	ns	ns	=[a]	=	=
Mean corrected change in trained stimulus	ns	ns	ns	Significant by selection			=	=	=
No. relation training trials	ns	ns	ns	ns	ns	ns	=	=	=
No. response training trials	ns	ns	ns	ns	ns	ns	=	=	=
Order of rating	See text			See text			Not tested		

[a] "=" means equated by experimental design.

stimulus), it is necessary to determine whether certain other variables were equated. Table 2 indicates five variables which were tested in order to determine whether they had significantly different values in the groups compared in the major analyses.

Time 1 ratings of stimuli.—The principal dependent variable of the experiment was the change in attitude scores determined by subtracting the Time 1 evaluative rating of each nonsense syllable from the Time 2 rating. It was immediately obvious that the uncorrected change score was influenced by the initial rating of a nonsense syllable. Those syllables which were initially positive were more likely to change in a negative direction, while the initially negative tended to become more positive at Time 2. The magnitude of this regression factor was estimated from the Test–Retest Control data which showed a correlation of $-.507$ between initial ratings and uncorrected change scores. Using a regression equation based on this correlation, all initial ratings of each *S* were converted to expected change scores. The result of subtracting each expected change score from the uncorrected change score was designated the *corrected change score*, with a negative score meaning that *S* changed her rating of the syllable more unfavorably than would have been predicted from knowing her initial rating. All change scores have been reported in this form unless otherwise indicated.

But differences in Time 1 scores might have other effects on attitude change which were not eliminated by the correction described. It is of importance, therefore, to examine the extent to which Time 1 ratings were equated. Row 1 of Table 2 shows that only two of the comparisons for untrained stimuli were significant; namely, the comparison of S_4 and S_7 in maximum change group and S_8 and S_7 in moderate change group. In Row 2 (trained stimuli) one was significant; the Time 1 rating of S_7 in the maximum change group was less positive than the Time 1 rating of S_7 in the moderate change group. One set of comparisons omitted from the Table also revealed a significant Time 1 difference: that between the rating of S_7 maximum and S_7 moderate under PA Control conditions. In other words, only four out of 108 comparisons involving Time 1 ratings of syllables in different serial positions, in different experimental conditions and at different levels of trained change were significant. Comment will be made in the discussion section on the possible relation of one of these significant differences to certain results.

Number of relations training trials.—For comparisons between serial positions within a trained change group the number of trials was necessarily the same. There were no significant differences between Serial and PA Control conditions within any level of trained change or between levels of trained change (see Row 4, Table 2).

Number of response training trials.—Row 5 of Table 2 shows that this variable was also equated in all the comparisons made.

Average rating order of stimuli.—Since the order of rating stimuli may influence judgments, it is important that this factor be equated for stimuli in different serial positions, in groups with different amounts of trained change and in Serial and PA Control groups. Because the number of cases within each cell was not large enough to make a satisfactory test of the difference in the distribution of rating positions for each condition, it has been necessary to determine whether the findings were changed when rating order was held constant. These analyses are presented later.

Change in Response to the Trained Stimulus

The total sample was used to test whether ratings of S_7, the stimulus associated with shock, became more negative. In both Serial and PA Control conditions the mean corrected change score of the trained stimulus was significantly more negative than was the mean corrected change score for the untrained stimuli. In the Serial condition all six comparisons had $Ps < .05$. In the PA Control condition the difference between the mean change in S_7 and the mean of individual average change scores of the six untrained stimuli was significant at the .001 level.

As already explained, it was expected that the amount of generalized attitude change would be a function of the amount of change in S_7 and the effects of serial position have been analyzed with change in S_7 held con-

TABLE 3

GENERALIZED CHANGE IN AFFECTIVE RESPONSES

Condition	Corrected Change Score for Trained Syllable											
	Minimum				Moderate				Maximum			
	N	Mean	*SD*	Diff.[a]	*N*	Mean	*SD*	Diff.[a]	*N*	Mean	*SD*	Diff.[a]
Serial												
S_4	16	0.9	13.0	−1.9	18	3.4	12.8	− 1.4	15	3.6	19.3	0.4
S_5	16	2.0	12.8	−0.8	18	6.2	15.7	1.4	15	3.4	17.8	0.2
S_6	16	2.0	11.5	−0.8	18	− 1.0	12.1	− 5.8*	15	17.0	11.9	13.8***
S_7 (Trained)	16	−7.0	2.5	−1.0	18	−14.0	1.8	0.0	15	−30.0	8.1	−5.0
S_8	16	5.8	13.6	3.0	18	− 6.8	16.4	−11.6**	14	− 5.0	14.6	−8.2*
S_9	16	0.4	12.1	−2.4	18	7.8	12.3	3.0	15	2.9	16.4	−0.3
S_{10}	16	4.6	12.1	1.8	18	9.0	13.2	4.2	15	0.6	15.9	−2.6
PA Control												
Untrained	11	2.8	4.1		13	4.8	4.9		15	3.2	7.1	
Trained	11	−6.0	2.6		13	−14.0	2.4		15	−25.0	9.6	

[a] Algebraic difference between Serial and PA Control; Untrained Serial compared with Untrained PA Control average, and Trained Serial compared with Trained PA Control.
* $P < .10$.
** $P < .02$.
*** $P < .01$.

stant at different levels (see Table 1). At none of the three levels of change in S_7 was there a significant difference between Serial and PA Control in the amount of change in the trained stimulus (Table 2, Row 3).

Generalization as a Function of Position within a Serial Structure

Change in attitude toward the untrained stimuli which were serially related to S_7 has been compared to the mean change of untrained stimuli in the PA Control group with level of trained change held constant. This comparison has also been made holding rating position constant at each level of trained change. Finally, change at each serial position has been compared with change at every other position, again with trained change held constant.

Comparison of change under Serial and PA Control conditions.—Hypotheses 2a, 3a, and 4a predicted that change in S_6 and S_8 would differ from change in untrained PA Control stimuli. Mean corrected change scores for Serial Positions S_4 through S_{10} in each of the three trained change groups are presented in Table 3 and Fig. 6. Entries for the untrained PA Control stimuli are means of individual average change scores for the six test syllables. The significance of differences between means for each list position and the mean of the untrained Control stimuli was tested at each level of trained change.[5]

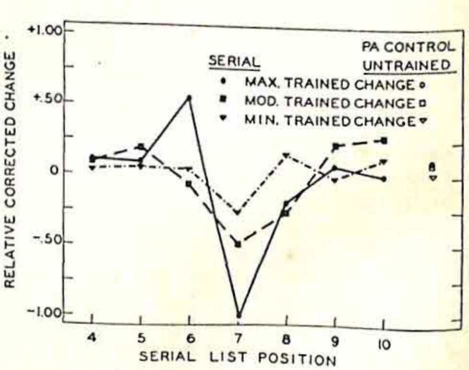


FIG. 6. Change in affective rating as a function of serial position. The ordinate is corrected change, divided by the absolute value of corrected change for S_7 in the maximum trained change group.

---

[5] Where estimated variances were heterogeneous, the approximation to *t* described by

For *S*s with maximum change in S_7 the rating of S_6 changed more positively than the average of PA Control ratings of all untrained test stimuli ($P < .01$)[6] and S_8 changed more negatively ($P < .10$). At the level of moderate trained change, ratings of S_6 tended to become more negative than PA Controls ($P < .10$), while S_8 scores were significantly more negative than PA Control scores ($P < .02$). There was no significant difference between any other serial Position and PA Control change score.

The results just reported have tested the hypotheses in question by comparing the amount of corrected change in attitudes from Time 1 to Time 2. However, a part of the theory from which predictions were derived involves assumptions about events at the time of the second rating of the stimuli. This means that the relations must have existed at Time 2. Since Time 1 ratings for the various stimuli and conditions were significantly different in only four of the 108 comparisons, it is reasonable to assume that differences in change scores will reflect differences in the Time 2 ratings of different stimuli. Nevertheless, this has been checked by computing mean Time 2 ratings for S_6, S_7, S_8 within each trained-change level and for the PA Controls trained and untrained within each level. The results obtained by this analysis were exactly parallel to those reported in terms of amount of trained change. The same comparisons were significant and the same ones fell below acceptable levels of significance.

Before stating the conclusions of these tests of the Hypotheses we turn, then, to one other check; namely, the effects of holding constant the order in which stimuli were rated.

Comparison of change under Serial and PA Control conditions with rating order constant.—It will be recalled that it was considered necessary to check the effects of our principal variables with order of rating the test stimuli held constant because of the impossibility of testing adequately whether this variable was equated. Table 4 presents results. They were essentially the same as those just reported.

This table was constructed in the following manner. Within a given level of trained change, mean change in S_6 was computed for Serial *S*s who rated this stimulus at each possible rating position (4, 6, 7, 8, 9, or 10). The average of PA Control test stimuli rated at each of these positions was also computed for the comparable trained change group. Using these values from Table 4, the PA Control mean at each rating position was subtracted from the Serial mean for S_6 in that same rating position. The same procedure was followed for S_8 change scores within each level

---

Dixon and Massey (1951, pp. 104–105) was used in testing significance. Since the measure of change for test syllables in the Controls was an average of six scores, its variance was lower than variance of the measures for the Serial condition. A two-tailed test was used throughout. A number of additional predicted relations would be significant at the .05 level if a one-tailed test were used.

[6] These results were based on all *S*s whose Time 2 ratings changed negatively, even though the hypotheses involving S_6 were not made explicit until one-third of the *S*s had been run. It should be noted, however, that S_6 results based *only* on the *S*s who were run following the revision of Hypotheses 3 and 4 were exactly parallel to those reported in the tables. The change in S_6 for the maximum change group was significantly ($P < .02$) more positive than the mean change of untrained stimuli in the PA Controls, maximum change group. Change in S_6 in moderate change group was more negative than for PA Control moderate change group ($P < .10$). For the minimum change group, $P > .10$.

of trained change. The significance of each set of six differences (where there were no empty cells) has been evaluated using a randomization test for matched pairs (Siegel, 1956, pp. 88–92).

With rating order held constant, the relations between corrected change scores of stimuli at different serial positions were maintained with surprising consistency considering the very small number of cases in each cell. As indicated in the Table, mean differences between Serial and PA Control conditions were significant for comparisons of S_6 and S_8 in the maximum and moderate change groups, respectively, but not significant ($P < .10$) for the other two comparisons.

The results strongly support Hypothesis 3a, while Hypotheses 2a and 4a received moderate support.

Comparison of change at different serial positions.—Hypotheses 2b, 3b, and 4b predict different amounts of generalization *within* a serial list. For each *S* in the Serial condition the difference between corrected change scores for each pair of list positions was computed and averaged over *S*s within each level of trained change.

The results in Table 5 indicate that when S_7 changed maximally, S_6 showed significantly more positive change than syllables at every other position, while S_8 was consistently more negative than all others except S_7 but significantly different only from S_6 and S_7.

At the level of moderate trained change S_6 changed more negatively than S_4, S_5, S_9, and S_{10}, the last three differences having *P*s of .10 or less; S_8 was more negative than every other test stimulus at this level and only in the comparison with S_6 and S_7 did the difference fail to reach the .05 level of significance. None of the comparisons involving pairs which did not include S_6, S_7, or S_8 were significant and no differences were significant in the minimum trained change group except those involving S_7.

TABLE 4

MEAN CORRECTED CHANGE FOR TEST SYLLABLES AT EACH RATING POSITION IN SERIAL AND PA CONTROL CONDITIONS

Trained Change Group	Test Syllable	Rating Position						Serial − PA Control	
		4	6	7	8	9	10	Mean Diff.	P*
Maximum	Serial S_6	31.0	9.0	20.0	20.0	12.0	22.0		
	PA Control	−5.2	4.4	10.3	1.0	3.5	5.5	15.7	<.04
	Serial S_8	6.7	−11.0	−11.5	−14.0	−14.0	2.7	−10.1	<.10
Moderate	Serial S_6	−1.5	−7.5	10.3	−9.0	7.0	−6.0		
	PA Control	−0.9	6.2	6.3	4.1	12.5	0.4	−5.9	<.10
	Serial S_8	−4.0	−13.8	−0.7	−4.7	−9.0	−1.0	−10.2	<.04
Minimum	Serial S_6	1.5	−1.0	[a]	2.2	10.0	1.7		
	PA Control	4.4	−0.7	1.3	7.1	−0.2	5.1	−0.3	ns
	Serial S_8	7.6	[a]	27.5	−7.5	0.5	2.8	2.6	ns

* By a randomization test on differences at rating positions.

[a] No cases in the Serial condition. This rating position was omitted in tests of significance.

TABLE 5

MEAN DIFFERENCE IN CORRECTED ATTITUDE CHANGE FOR PAIRS OF SERIAL POSITIONS

Serial Positions	Trained Change Group								
	Maximum			Moderate			Minimum		
	N	M	SD	N	M	SD	N	M	SD
S_7-S_4	15	−33.3****	23.3	18	−17.4****	12.4	16	− 7.6**	12.0
S_7-S_5	15	−33.1****	18.2	18	−20.2****	15.0	16	− 8.8**	12.2
S_7-S_6	15	−46.7****	15.0	18	−13.0****	12.2	16	− 8.8**	11.7
S_7-S_8	14	−24.1****	11.2	18	− 7.2	17.4	16	−12.6****	14.3
S_7-S_9	15	−32.6****	18.1	18	−21.8****	12.1	16	− 7.2**	11.8
S_7-S_{10}	15	−30.3****	15.8	18	−23.1****	13.5	16	−11.3****	12.4
S_6-S_4	15	13.4**	23.2	18	− 4.4	14.7	16	1.1	14.2
S_6-S_5	15	13.6**	21.2	18	− 7.2*	15.6	16	0.0	16.2
S_6-S_8	14	20.7****	17.3	18	5.8	17.4	16	− 3.8	21.2
S_6-S_9	15	14.1****	17.3	18	− 8.8*	17.6	16	1.6	11.2
S_6-S_{10}	15	16.4****	19.5	18	−10.0**	18.7	16	− 2.6	13.8
S_8-S_4	14	− 7.8	25.4	18	−10.2****	14.9	16	4.9	20.7
S_8-S_5	14	− 8.4	21.2	18	−13.0***	24.3	16	3.8	22.7
S_8-S_9	14	− 5.9	17.2	18	−14.6***	23.6	16	5.4	20.1
S_8-S_{10}	14	− 5.4	17.6	18	−15.8****	20.9	16	1.2	18.4
S_4-S_5	15	0.2	32.7	18	− 2.8	19.4	16	− 1.1	13.2
S_4-S_9	15	0.7	21.3	18	− 4.4	17.2	16	0.4	14.7
S_4-S_{10}	15	3.0	17.4	18	− 5.6	17.8	16	− 3.7	18.1
S_5-S_9	15	0.5	27.9	18	− 1.6	20.1	16	1.6	15.3
S_5-S_{10}	15	2.8	27.7	18	− 2.8	23.8	16	− 2.6	19.3
S_9-S_{10}	15	2.3	17.6	18	− 1.2	20.8	16	− 4.1	12.0

* $P < .10$.
** $P < .05$.
*** $P < .02$.
**** $P < .01$.

The results parallel those of the previous analysis: Hypothesis 3b was strongly supported while 2b and 4b received support at a borderline level. In all analyses, S_8 in the moderate trained change group was significantly negative as had been predicted for S_8 in the maximum trained change group, while in the latter group there was less significant negative change.

Generalization as a Function of Change in the Trained Stimulus

Hypotheses 1 and 3c predict that amount of generalization within a serial structure is a function of level of change in response to the trained stimulus. A direct comparison of Serial and PA conditions at different levels of trained change could not be made because of inhomogeneity of variance. In the PA condition, however, differences between change levels in average corrected change scores for the untrained stimuli were small and not significant.

Interlevel comparisons of change in S_6 and S_8.—The comparison of mean corrected change for corresponding serial positions at the three levels of trained change (Table 3) indicated that change in S_6 was significantly more positive when S_7 changed maximally than in either the minimum or moderate change conditions ($P < .005$ in both cases). Change in S_6 was not significantly different between the latter two levels.

The same comparisons for S_8 showed significantly more negative change in the moderate than the minimum condition ($P < .05$) but the predicted difference in the same direction between the maximum and minimum change groups did not reach an acceptable level of significance with a two-tailed test ($P < .10$). The difference between the maximum and moderate change groups was opposite that predicted and was not significant.

Interlevel comparisons of change in S_6 and S_8 with rating order held constant.—Differences between levels of trained change in mean S_6 scores at each rating position (computed from Table 4) were evaluated using the randomization test for matched pairs (Siegel, 1956, pp. 88–92). The same procedure was followed for S_8. By this analysis all differences in S_6 change between levels of trained change were significant or near significant (maximum>moderate, $P > .04$; maximum > minimum, $P < .07$; and moderate<minimum, $P < .07$). Only one difference for S_8 approached significance (maximum < minimum, $P < .07$). The symbol (<) means more negative; (>) more positive. These results differ somewhat from those where rating order was not controlled. The near significant difference for S_6 between moderate and minimum change groups was not observed there, and this is the only analysis where the significance of change for S_8 was greater in the maximum than in the moderate trained change groups.

Interlevel comparison of difference between serial positions in attitude change.—Essentially the same results were obtained when difference between change in ratings of S_6, S_8, and other untrained list syllables was investigated as a function of change in S_7 (Table 5). The mean difference between S_6 and each other serial position at the maximum level of change in S_7 was subtracted from the mean difference between S_6 and that same position at a different level of trained change, and the significance of the difference determined. This comparison was made between levels for every difference involving S_6 and S_8.

All five differences involving S_6 were significant ($P < .05$) in comparisons of maximum and moderate levels of trained change; in maximum–minimum comparisons three were significant at the .05 level and one at the .10 level; one was significant at the .10 level in comparison of moderate and minimum trained change groups.

For S_8 only comparisons between moderate and minimum levels of trained change were significant (four out of five at the .05 level). In all comparisons of the maximum and minimum groups the differences were in the expected direction: S_8 was relatively more negative than other positions in the maximum than in the minimum group. Here, as in all but one other analysis, the greater amount of negative change in the moderate than in the maximum change group was apparent, though the differences were not significant.

It may be concluded from all methods of analysis that there was strong support for Hypothesis 3c in the greater positive change in S_6 when S_7 changed maximally. Hypothesis 1 was supported by the fact that greater negative change in S_8 occurred when change in S_7 was moderate than when it was minimum. However, the general finding of slightly more negative change of S_8 in the moderate than in the maximum trained group was not predicted.

Discussion

The hypotheses have been supported in general except for 1 and 2a, which predicted the greatest negative change in S_8 when S_7 changed maximally. Actually S_8 changed more negatively in the moderate group than in the maximum change group. Two possible reasons for this result suggest themselves. First, there is a possibility of the operation of an artifact incident to the significantly more positive Time 1 rating of S_7 in the maximum change group than in the moderate group in both Serial and PA Control conditions. The question is whether this different Time 1 rating has made the amount of trained change an appropriate basis for choosing groups since under these circumstances the Time 2 rating of S_7 in the moderate change group *might have been* more negative than the Time 2 rating of the maximum change group. If this were so, then it would be expected that the Time 2 rating of S_8 moderate might be more negative than the rating of S_8 maximum. A check on the mean Time 2 ratings of S_7 in each group indicates that this was not the case. The Time 2 ratings in the maximum, moderate, and minimum trained change groups had the same order indicated by amount of trained change; moreover, the differences in Time 2 scores for the three groups were highly significant, all with probabilities less than .001.

Still another hypothesis is plausible. It will be recalled that our prediction about the effects of events occurring at the time of testing took into account the possible influence only of the immediate forward associate of the stimulus being tested. This was recognized as an oversimplification which, as it now appears, may need reconsideration. It may be that *multiple* events elicited by a stimulus combine in various ways to influence the rating which *S* gives to a stimulus. Only two possible loci of such a combination can be mentioned at this time, both of which should have similar effects. First, assume that *either* the immediate forward or the immediate backward associate is elicited on any one occasion with the probabilities suggested by the Raskin and Cook data. Then the forward associate will provide the context of rating a given stimulus on some proportion of the semantic differential scales and the backward associate will be the context for the ratings on other scales (five evaluative scales were checked by each *S* for each stimulus). Or, secondly, it may be that *each* judgment on a single scale is an average or some combination of interactions between $r_{a,n}$ and $r_{a,n+1}$ and interactions between $r_{a,n}$ and $r_{a,n-1}$ all of which are involved in the process of judging the position of S_n on one scale. Again, the frequencies of each of these two types of interaction would be proportional to the probabilities of the elicitation of a forward and a backward associate, respectively.

Take as an example the situation under discussion, the rating of S_8 in the moderate and the maximum change groups. Suppose that S_8 when presented for rating elicits $r_{s,8}$ which in turn elicits $r_{s,9}$ and/or $r_{s,7}$ in the ratio of about 30/7 which corresponds to the relative frequency of occurrence of immediate forward and backward associates. The suggestion is that the judgment of S_8 would be weighted more heavily by the results of interaction of $r_{a,8}$ with $r_{a,9}$ than by the interaction of $r_{a,8}$ with $r_{a,7}$ (whether the average is over scales or over events involved in a single judgment or over individuals). Now, since the $r_{a,7} - r_{a,8}$ distance would more likely reach contrast levels when S_7 change is maximal, this would be the condition under which the judgment of S_8 would be most influenced in the *positive* direction by the elicitation of $r_{a,7}$. On the other hand, at moderate change levels elicitation of $r_{a,7}$ would more probably lead to assimilation of $r_{a,8}$ to $r_{a,7}$ which would tend to make S_8 more *negative*. In short, under conditions of maximum change in S_7 the contribution of the backward associate to the judgment of S_8 would be to make it more positive and under conditions of moderate trained change the contribution of the elicited backward associate would be to make the judgment of S_8 more negative.

This *post hoc* suggestion about the possible effects of backward associates occurring at the time of testing for generalization serves to introduce a final comment about the complexities of the problem of generalization. We have considered *one* structural pattern (serial associates), *one* aspect of structural change taking place as a result of response training, and the interaction of *one* set of events at the time of testing: the affective response associated with the test stimulus at the time of response training, the forward associate of the test stimulus and its attached affect. It is at once evident that change in conditions at each of these three points will alter the response made to any stimulus tested for generalization. First, there are other structures within which generalization occurs and these will differ in the number of related parts and in the probabilities of arousal of each part by every other part.

Secondly, at the time of response training the structure may be changed in a variety of ways by different input patterns operating on initially different structures. To take but one example: we have some evidence (preliminary and unpublished) that when a positive affective response is attached to one point in a serial structure and negative affect to another point in the same structure, the serial relation no longer plays the dominant role in determining the occurrence of generalization from one point to another within the serial structure. Finally, the response elicited at the time of testing for generalization may be altered by the interaction of a variety of concurrent events initiated by the particular instructions, by the pattern of contextual stimuli, and by the whole matrix of associations that may be activated by the stimulus-to-be-tested as a result of existing structural arrangements.

A concluding comment must be made regarding the relation of the theory here presented to the more conventional mediation theory. Although the results reported at this time are not those that would have been predicted by the mediation theorists, there is no doubt that under some conditions presentation of an untrained stimulus will have some probability of being followed by the response which has been associated with the forward associate of the test stimulus and this may occur because at the time of testing the untrained stimulus elicits the "mediating" forward associate which in turn elicits the response. But neither is there any doubt that generalization may appear under other conditions. It is our suggestion that a complete theory of generalization must specify the different conditions under which generalization occurs. And this may be accomplished only by taking into account the structural arrangements and the input patterns operating both at the time of training the response and of testing the untrained stimuli for evidences of generalization.

SUMMARY

A theory is presented which attributes the nature of the response made to a stimulus being tested for generalization within a serial structure to the combined effects of (*a*) associations formed at the time that the stimulus-in-training is being associated with a response, and (*b*) the interaction at the time of testing between the response to the test stimulus and responses to other events associated with the test stimulus.

The *S*s rated 12 nonsense syllables on evaluative scales. One group then learned the nonsense syllables in serial order after which the seventh syllable was associated with shock. Syllables were rated a second time. A PA Control group followed the same procedures except that in place of serial training each of the 12 syllables was the response to two-place number in paired-associates training. An analysis was made of changes in the evaluative ratings of these stimuli for *S*s in three groups: those for whom the change in the trained stimulus was maximal; a moderate trained change group; and a minimum trained change group.

The following results were obtained:

1. Trained stimuli changed significantly in both the Serial and PA Control conditions when testing each group as a whole.

2. The evaluative ratings of S_8 immediately following the trained stimulus changed significantly more in the negative direction than did comparable PA Controls and more than serial stimuli in other list positions, when S_7 changed moderately. The negative change in S_8 in the maximum change group was of smaller magnitude than in the moderate trained change group. This was contrary to predictions.

3. In the maximum trained change group the stimulus immediately preceding the trained stimulus changed more positively than any other serial stimuli and more positively than comparable PA Controls, as predicted. This positive change in S_6 disappeared as the change in the trained stimulus decreased and tended to be negative at the moderate level of trained change. This was expected.

It is proposed that the positive change in S_6 at maximum trained change levels resulted from contrast with the very negative implicit response to S_7 which was elicited at the time of testing S_6. The negative change in S_8 was predicted on the basis of events assumed to take place at the time of response training.

REFERENCES

BUGELSKI, B. R. A remote association explanation of the relative difficulty of learning nonsense syllables in a serial list. *J. exp. Psychol.*, 1950, **40**, 336–348.

DIXON, W. J., & MASSEY, F. J. *Statistical analysis.* New York: McGraw-Hill, 1951.

MINK, W. D. Semantic generalization as related to word association. Unpublished doctoral dissertation, Univer. of Minnesota, 1957.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *Measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PEAK, H. Psychological structure and person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. (a)

PEAK, H. Psychological structure and psychological activity. *Psychol. Rev.*, 1958, **65**, 325–347. (b)

RASKIN, E., & COOK, S. W. The strength and direction of associations formed in the learning of nonsense syllables. *J. exp. Psychol.*, 1937, **20**, 381–395.

RAZRAN, G. H. S. Semantic and phonetographic generalizations as a function of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, **39**, 642–652.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

(Received May 27, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

TRANSFER OF A DISCRIMINATION FROM OBJECTS TO PATTERNS[1]

BETTY J. HOUSE AND DAVID ZEAMAN

University of Connecticut

Lawrence (1952) presented evidence for an empirical law of discrimination learning that, in training a hard discrimination, it is more efficient to train first on easier discriminations along the same dimension than to give all trials on the hard discrimination. After equal numbers of trials, rats with all training on a difficult brightness discrimination performed more poorly than those having trials first on easier brightness discriminations. Baker and Osgood (1954) observed similar effects with adult humans in pitch discrimination. These results provide an exception to the usual finding of transfer studies that maximum performance is obtained when training and test conditions are identical. In the present study, an attempt is made to extend the generality of the law by testing it with a different kind of *S*, mentally retarded children, and a different type of problem, color-form discrimination.

A difficult discrimination problem for mentally retarded *S*s is one in which the stimuli are patterns painted on a background. An easier problem can be obtained by presenting stimuli as three-dimensional objects. Even in the case where the relevant cues are identical, an object discrimination is learned more quickly than a pattern discrimination by retarded *S*s (House, Zeaman, Orlando, & Fischer, 1957) as well as by monkeys (Harlow, 1945). Thus, preconditions for a test of the law demonstrated by Lawrence are present: an easy and a hard discrimination problem with the same relevant dimension.

It might be questioned how object and pattern stimuli can be ordered along the same stimulus dimension, as required for application of Lawrence's law. In answer, the two object stimuli may be regarded as differing along a continuum of partial identity. The positive and negative objects are composed of some common stimulus elements as well as distinctive relevant cues. The present pattern stimuli are conceived as having the same relevant elements as the objects, but with an added number of common background elements. Thus the pattern stimuli occupy intermediate positions between the two objects along a continuum of partial identity, conceivably measurable in terms of overlap of sets of elements.

In this experiment, discrimination was trained between stimuli differing in color and form, presented in either of two ways: as flat patterns on gray backgrounds or as cut-out three-dimensional objects. One group (P-1) was given trials on the pattern stimuli only. Another group (0–1 to P–1) was trained first on a pair of object stimuli before being shifted to a pair of pattern stimuli having the same colors and forms as the

---

[1] This research was carried out at the Mansfield State Training School, Mansfield Depot, Connecticut. The authors wish to express appreciation to Neil A. Dayton, Superintendent, and to John Cassell and Louis Boly, principal and assistant principal of the Longley School, for their aid and cooperation. This investigation was supported by Research Grant M-1099 from the National Institute of Mental Health.

TABLE 1

SUBJECT CHARACTERISTICS OF THE EXPERIMENTAL GROUPS

Condition	*N*	CA (Mo.)		IQ		MA (Mo.)	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
0–1:P–1	19	128	34	39	12	47	12
0–2:P–1	22	136	35	41	11	53	13
P–1	19	132	40	40	13	49	12

objects. This condition is analogous to Lawrence's easy-to-hard group. A third group (0–2 to P–1) was trained on object stimuli different in color and form from the pattern stimuli. In this condition, approach-avoidance tendencies acquired to the object stimuli cannot be directly transferred to the new pattern stimuli.

METHOD

Apparatus

The discrimination apparatus consisted of a table with a one-way screen separating *E* from *S*, and a sliding tray for presentation of the stimuli. Two 3-in. foodcups were set into the 30-in. tray, centered 12 in. apart. The tray was pulled out of sight behind the one-way screen for baiting and pushed diectly in front of *S* to begin a trial.

Two pairs of color-form stimuli were constructed in the form of object stimuli and pattern stimuli. These were (*a*) yellow T vs. black square and (*b*) red cross vs. green circle. For the object stimuli, forms were cut from $\frac{1}{4}$-in. Masonite and mounted vertically on 4-in. × 4-in. gray Masonite bases. For the pattern stimuli, forms of the exact dimensions and colors of the object stimuli were painted on gray wedges with 4-in. × 4-in. faces. Each form had maximum height and width of 2 in.

Subjects

All *S*s were experimentally naive, selected randomly from boys and girls attending classes at the Longley School of the Mansfield State Training School, within the mental age range from 2 yr. to 6 yr. The *S*s were assigned to the three groups so as to equate for CA, IQ, and MA, as shown in Table 1.

Procedure

General.—The *S*'s response was lifting the stimulus. Under the correct stimulus, *E* placed a small piece of candy. Reinforcements were M & M's, candy corn, Malties, and the like, which *S* could eat or save as he preferred. As additional reinforcement, *E* said "Good" if the response was correct and "No" if it was not.

One of the two pairs of stimuli was assigned randomly to each *S* and the positive cue was designated randomly. The correct stimulus was placed on the left or right according to a Gellerman's series (1933). All *S*s were given 25 trials per day with correction procedure. The criterion for learning was 20 out of 25 trials correct during a single daily session.

Pretraining.—Verbal instructions were limited to asking *S* to find the candy. Before the first experimental session, several trials were given in which *E* placed the candy in an open foodcup and then under a single gray wedge placed randomly left or right.

Experimental conditions.—The sequences of training trials for the three groups were as follows. In Group 0–1:P–1 there was object training to criterion with a maximum of 250 trials, followed by identical pattern training to criterion or 250 trials. In Group 0–2:P–1, there was object training to criterion with a maximum of 250 trials, followed by different pattern training to criterion or 250 trials. In Group P–1 there was pattern training to criterion with a maximum of 500 trials.

RESULTS

Performances of the three groups are shown in Fig. 1. The percentage of *S*s who have met criterion on the *pattern* problem is plotted as a function of total training days. For Group 0–1:P–1 and Group 0–2:P–1, total training days refers to the sum of days spent on the object problem *and* the pattern problem. For example, the percentage on Day 2 for Group 0–1:P–1 signifies that 21% learned 0–1 to criterion on the first day and learned P–1 to criterion on the second day. The superiority of the two object-trained groups over Group P–1 is significant. Chi square tests were made of the frequencies of

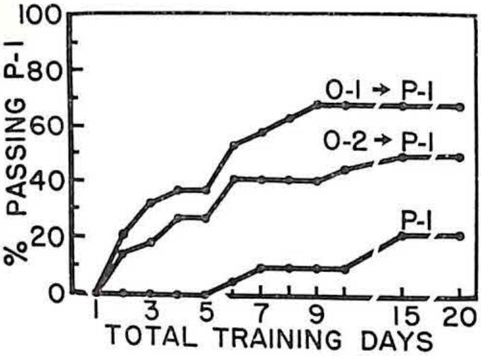


FIG. 1. Percentage of *S*s in each group who have learned pattern is plotted as a function of days of training. For Group 0–1:P–1 and Group 0–2:P–1, total days refers to sum of object and pattern training days.

*S*s who had passed or had not passed the pattern test by the end of 10 days training. The probability that differences as large as those observed would occur by chance is less than .001 for Group 0–1:P–1 and less than .02 for Group 0–2:P–1. Little change occurs during the final 10 days during which all *S*s who had failed object or pattern were trained on pattern.

While both kinds of object training produced significant facilitation of pattern learning, the question remains whether there is an added advantage when the relevant cues are identical for the object and pattern problems. A chi square test failed to show a difference between percentage passing pattern for the two object groups ($P > .10$). However, when a more sensitive measure was considered, that of number of errors to criterion on the pattern problem, Group 0–1:P–1 was superior ($P < .01$, Mann-Whitney *U* test, one-tailed). Mean errors were 5 for Group 0–1:P–1 and 47 for Group 0–2:P–1; median errors were 2 and 9, respectively. That the better performance on pattern was not due to chance differences in learning ability is demonstrated by the fact that the two groups were equivalent in performance on the object problem ($P > .70$, Mann-Whitney *U* test). Percentages of *S*s learning the object problem were 68 and 63 for 0–1 and 0–2, respectively. All *S*s in Group 0–1:P–1 who learned the object discrimination also learned the pattern discrimination, while of the 14 *S*s who learned the 0–2 problem, only 10 then learned the pattern problem.

It is recognized that the curves of Fig. 1 do not reveal the shapes of the underlying learning functions. These are of interest in relation to another theory and will be discussed in detail elsewhere. All the individual functions can be classified as belonging to a family of ogival functions which pass from chance to criterion very quickly, once the curves start to rise. Wide individual differences exist, however, in the duration of the chance stage of performance. Therefore, since *S*s are for the most part either at chance or criterion, the reported measure of percentage passing is a good approximate description of the state of learning of each group.

DISCUSSION

Our results support Lawrence's conclusion that a training sequence of easy-to-hard discrimination trials with the same relevant dimension is more efficient than a training series of hard discrimination trials only. In addition, our 0–2:P–1 condition demonstrated that direct transfer of approach-avoidance tendencies is not necessary for a facilitating effect. This finding adds weight to Lawrence's (1955) argument that current theories of generalization gradients cannot account for his results. If there is generalization due to the similarity of the 0–2:P–1 stimuli, the effect is as likely to be interference as facilitation since the positive cue is randomly designated. While generalization of approach-avoidance tendencies from the 0–1 training to the P–1 condition presumably accounts for the superiority of 0–1 training over 0–2 training, there remains a large facilitating effect which requires some other explanation.

A model for discrimination learning developed by Restle (1955), which relates learning rate to proportion of relevant and irrelevant cues, has been shown to account for Lawrence's results quite parsimoniously. This theory can be applied to our data to predict the amount of 0–1:P–1 transfer from a knowledge of the learning rate parameters (θ) of the object and pattern learning curves. Parameters were estimated for the four *S*s (i.e., 21%) of Group P–1 who learned the pattern discrimination and for the comparable top 21% (4 *S*s) of Group 0–1:P–1. With mean θs equal to .27 and .02 for the object and pattern *S*s, respectively, Restle's Equation 10 (1955, p. 15) was used to calculate predicted cumulative errors for the four 0–1:P–1 *S*s on the first 25 transfer trials. Table 2 compares the obtained errors with the predicted errors. Although the number of *S*s in this comparison is small, the magnitude of the difference between predicted and obtained errors appears sufficient to warrant rejection of Restle's transfer Equation 10 for these data.

It should be pointed out that Restle's Equation 10 requires the assumption that object and pattern stimuli have the same *relevant* cues, but that pattern stimuli have more *irrelevant* cues than appear in the object stimuli. An alternative assumption could be made: objects have more relevant cues than pattern, but the same irrelevant cues. Equations may then be derived from Restle's theory which predict the almost perfect transfer found in this experiment. Thus, if Restle's theory is applicable in the present experiment, the implication is that in going from 0–1 to P–1 no irrelevant cues are added, but relevant cues are subtracted. The theory makes no predictions about transfer from 0–2 to P–1.

TABLE 2

PREDICTION OF TRANSFER FROM 0–1 TO P–1

Transfer Trials	Cumulative Errors Observed	Cumulative Errors Predicted
5	2	9
10	2	18
15	2	27
20	2	36
25	3	45

Lawrence (1952) suggested that his results may be due to acquired distinctiveness of the relevant stimulus dimension. Stimuli acquire distinctiveness during easy discrimination training when *S* learns to "isolate functionally the relevant stimulus dimensions from all other background and irrelevant cues" (Lawrence, 1952, p. 516). The distinctiveness then transfers to the difficult discrimination. Details of the process are not elaborated by Lawrence. However, a more detailed analysis of learning of this sort has been developed by Wyckoff (1952), whose model for "observing responses" may account for our results. Wyckoff defines the "observing response" (R_0) as any response which exposes *S* to the relevant cue. For our experiment, we might interpret R_0 as *looking at the color and form cues.* The probability (p_0) of R_0 has been postulated to bear a circular relation to the speed of discrimination learning. Each is dependent upon and facilitates the other. Response R_0 is learned to the extent that its occurrence increases the probability of reinforcement, but in a discrimination problem R_0 can increase probability of reinforcement only if *S* is responding above chance, i.e., is learning.

This reasoning may be applied to our data. For object stimuli, p_0 may be initially higher than for pattern, because of fewer competing background (non-color/form) cues. This condition would facilitate the object discrimination and in turn increase p_0 even further. Consequently *S*s going from object to pattern would have a special advantage in starting pattern with p_0 higher than that ever developed by Group P–1 during its unsuccessful pattern training.

While Wyckoff's theory may thus be capable of accounting for our nonspecific type of object-to-pattern transfer, his formal model was not designed for the particular situation of the present experiment. A modification of Wyckoff's model is now in preparation which we hope will also handle the quantitative aspects of the data.

SUMMARY

Two different training sequences of easy-to-hard discrimination trials were compared for efficiency with a training sequence of hard discrimination trials only. The *S*s were three independent groups of mentally retarded children equated for MA, IQ, and CA. The hard-discrimination group was trained with pattern stimuli differing in color and form. The two easy-to-hard groups were first trained to criterion with a pair of object stimuli before pattern training. Object training differed for the two groups only in that one group had identical relevant cues for the object and pattern problems while the other group had different colors and forms for object and pattern. Both easy-to-hard sequences were more efficient than the series of hard-discrimination trials only, as measured by total trials required to learn the hard discrimination. The presence of identical relevant cues in object and pattern increased the facilitating effect of object training. Within 10 training days, 68% learned both the object problem and the equivalent pattern problem, while in the other easy-to-hard group, 45% learned an object problem and a different pattern problem. In contrast, only 10% of the group having all training trials with pattern stimuli learned the pattern problem within 10 days. Implications of these data for Restle's discrimination model and Wyckoff's theory of "observing responses" were discussed.

REFERENCES

BAKER, R. A., & OSGOOD, S. W. Discrimination transfer along a pitch continuum. *J. exp. Psychol.*, 1954, **48**, 241–246.

GELLERMAN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 206–208.

HARLOW, H. F. Studies in discrimination learning in monkeys: V. Initial performance by experimentally naive monkeys on stimulus-object and pattern discrimination. *J. gen. Psychol.*, 1945, **33**, 3–10.

HOUSE, B. J., ZEAMAN, D., ORLANDO, R., & FISCHER, W. *Learning and transfer in mental defectives. Progress rep. No. 1*, NIMH, USPHS 1957. Res. Grant M-1099 to Univer. of Connecticut. Pp. 58–63.

LAWRENCE, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.

LAWRENCE, D. H. The applicability of generalization gradients to the transfer of a discrimination. *J. gen. Psychol.*, 1955, **52**, 37–48.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

WYCKOFF, L. B., JR. The role of observing responses in discrimination learning. Part 1. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received May 28, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

EYELID CONDITIONING AS A FUNCTION OF INTENSITY OF CONDITIONED AND UNCONDITIONED STIMULI[1]

EVELYN G. WALKER

University of Illinois

Although conditioned stimulus (CS) intensity has theoretically been considered one of the variables determining the strength of conditioned responses (CRs), the experimental evidence is in conflict. While some studies (Barnes, 1956; Brown, 1942; Hovland, 1937; Hull, 1949; Razran, 1957) have shown response strength to increase with increasing intensities of CS, others (Carter, 1941; Grant & Schneider, 1948, 1949; Taylor, 1954) have failed to find any relationship. Those experiments in which response strength was measured during acquisition of the CR typically were successful in demonstrating a positive relationship between CS intensity and strength of response. However, in general, those experiments in which the analysis was based on extinction data have reported negative findings. Within a single experiment (Kessen, 1953), the positive relationship shown for acquisition of a CR could not be demonstrated during extinction. Since the one important difference between acquisition and extinction of a CR is the omission of the unconditioned stimulus (UCS), it may be argued that this factor could, in part, have contributed to the conflicting results. If this is the case, then the effectiveness of CS intensity on response strength may depend upon the strength of the UCS.

One purpose of the present study was to determine whether UCS intensity is one of the conditions which influences the relationship between CS intensity and CR strength. Specifically, the hypothesis was tested that CS intensity would have a greater effect on response strength under a strong UCS than under a weak UCS. Such a result may be deduced from Hull's assumption of a multiplicative relation between habit, drive and stimulus intensity dynamism (V) (Hull, 1951), if UCS magnitude is viewed as a determinant of drive strength. Indirect support for this hypothesis may be found in the work of Pavlov (1927), who found that unless the UCS were strong, ordinarily suitable stimuli were ineffective as CS. In a recent analysis of studies on the conditioned salivary response, Razran (1957) has indicated that unconditioned response (UCR) magnitude is one of the variables related to the efficacy of CS intensity. He reported that the higher the UCR magnitude, the greater the effect of CS intensity on response strength. Since magnitude of the UCR is, in part, a function of UCS intensity, this could be interpreted as indirect evidence that the effect of CS intensity on response strength is related to the level of the UCS. Although in three studies with the eyelid response (Carter, 1941; Grant & Schneider, 1948; Taylor, 1954) no reliable relationship between CS intensity and strength of conditioning could be demonstrated, the intensity of the UCS in these studies would appear to have been relatively weak.

---

[1] This paper is based on a portion of a dissertation submitted to the Graduate College of the University of Illinois in partial fulfillment of the requirements for the PhD degree. The writer wishes to express her appreciation to G. Robert Grice, who directed the study, for valuable aid and advice offered throughout the course of the experiment.

If the effect of CS intensity on strength of response is altered by variations in the intensity of the UCS, then it would be difficult to demonstrate such a relationship using the typical extinction procedure where the UCS is omitted. To avoid this problem, the UCS was presented on each trial during extinction in the present experiment but with a CS-UCS interval which has been shown to produce extinction (McAllister, 1953).

A further purpose of the present study was to determine whether CS intensity has an effect on learning as distinct from performance. In his stimulus intensity dynamism principle, Hull (1951) has postulated that both learning (habit strength) and performance (reaction potential) are direct functions of CS intensity. Three experiments (Grant & Schneider, 1948, 1949; Kessen, 1953) have yielded no evidence to support the hypothesis that CS intensity affects habit strength. However, in these studies CS intensity did not influence response strength, and under conditions which fail to show this relationship, there would be little reason to expect an effect of CS intensity on habit strength.

METHOD

Subjects and design.—The *S*s were 160 men students of introductory psychology at the University of Illinois. Twenty *S*s were assigned at random to each of the eight cells at a 2 × 2 × 2 factorial design. During acquisition, which consisted of 80 paired presentations of tone and air puff separated by 500 msec., there were four different stimulus conditions: weak CS–weak UCS; strong CS–weak UCS; weak CS–strong UCS; strong CS–strong UCS; with two groups under each condition. There were eight conditions during extinction with all groups receiving the same UCS intensity as during acquisition. Four groups received the same CS intensity as acquisition. Two groups were shifted from weak to strong CS intensity, and two groups were shifted from strong to weak CS intensity. For extinction all groups received 30 trials with an interstimulus interval of 2500 msec., the air puff being presented on all trials.

Apparatus.—The CS was a tone of 1000 cycles, with the weak intensity 30 db, and the strong intensity 80 db above an average threshold. The tone was generated by a Hewlett-Packard, Model 205 AG audio oscillator and was delivered to *S* through a set of headphones.

Two intensities of air puff were employed as UCS; 0.5 lb./sq. in. in the case of the weak puff, and 5.0 lb./sq. in. in the case of the strong puff. Appropriate pressures were obtained by the use of a variable air pressure reducing valve. The CS had a duration of 500 msec. during acquisition and 2500 msec. during extinction, with the UCS onset coinciding with the termination of the CS and lasting 40 msec. A polygraph recorded the onset and offset of the stimuli which were controlled by means of Hunter electronic timers. The potentiometer-polygraph system used for recording eyelid movement has been described in detail by Spence (1953).

Conditioning procedure.—The *S*s were seated in a soundproof room. Incident illumination measured at the position of *S*'s eyes was approximately 2.5 ft-c. The ready signal was a weak buzzer which was presented at random intervals of 2, 3, or 4 sec. before each trial. At the ready signal, *S* was instructed to fixate a white disc (located 60 in. from *S* and 3 in. above eye level) and was told to blink once. An average intertrial interval of 20 sec. was used, with trials randomly separated by 15, 20, or 25 sec.

Measurement of CR.—Any response which occurred within 150 msec. of the CS onset was not counted as a CR since this is the interval in which original or unconditioned responses to the onset of the tone typically occur. A movement of the penwriter 1 mm. or more in amplitude in the interval 150 to 565 msec. after the onset of the CS was considered to be a CR. Although the UCS onset was 500 msec. after the onset of the tone, responses which occurred up to 565 msec. after CS onset were included as CRs when it was empirically determined that the minimum latency of the UCR was 75 msec.

In a previous study (Spence, 1953)

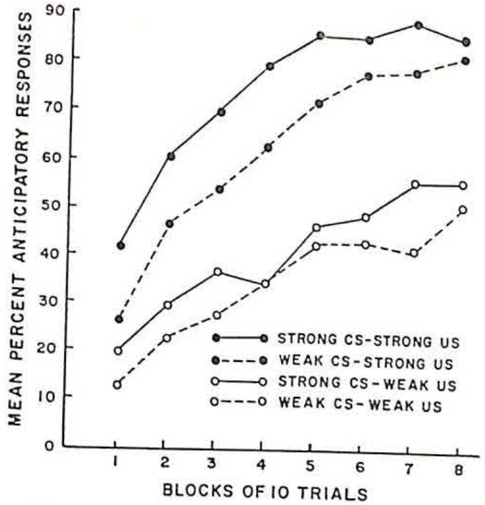


FIG. 1. Mean percentage of anticipatory responses in the interval 150 to 565 msec. during acquisition per block of 10 trials for the four stimulus conditions.

anticipatory responses were differentiated in terms of their latency, with the shorter latency responses being regarded as "voluntary" responses. Under the conditions of the present experiment, responses which occurred in the interval 150 to 300 msec. after CS onset were regarded as short latency or "voluntary" responses. Any *S* for whom more than 50% of his responses were in this interval was discarded as a voluntary responder. Sixteen *S*s were eliminated for this reason and one additional *S* was discarded because of failure to condition, the conditioning criterion being at least one CR during the 80 trials.

During extinction, CRs were also counted in the interval from 150 to 2500 msec. which allowed the inclusion of CRs whose latency had increased beyond the shorter interval (McAllister, 1953).

RESULTS

Acquisition of anticipatory response.—Figure 1 shows the mean percentage of anticipatory responses plotted as a function of blocks of 10 trials. For this analysis, an anticipatory response was defined as one whose latency was in the interval from 150 to 565 msec. after CS onset. Ordering of the groups is clear, the strong UCS groups showing consistently superior performance to the weak UCS groups, and the strong CS intensity groups exceeding the weak CS intensity groups.

A factorial analysis of variance was performed in order to evaluate the reliability of the observed differences in rate of acquisition of anticipatory responses under differential conditioned and unconditioned stimulus intensity. Table 1 summarizes the results of the analysis. The assumption of homogeneity of variance appeared tenable since Bartlett's test resulted in a chi square value which was not significant. It can be seen from Table 1 that the analysis yielded *F* values significant beyond the .01 level for both CS and UCS intensity, with no significant interaction. The test for interaction is a test of the hypothesis that CS intensity has the same effect on performance whether the UCS be strong or weak. This hypothesis cannot be rejected. Although it can be seen from Fig. 1 that the difference in performance under strong and weak CS was larger under the strong UCS than the weak UCS, this difference was not large enough relative to error variance to be reflected in

TABLE 1

ANALYSES OF VARIANCE OF TOTAL NUMBER OF ANTICIPATORY RESPONSES DURING ACQUISITION FOR THE EIGHT GROUPS ACCORDING TO LATENCY OF RESPONSE

Source	*df*	150–300 Msec. *F*	300–565 Msec. *F*	150–565 Msec. *F*
CS Intensity (CS)	1	13.89*	1.52	7.15*
Replications (R)	1	0.17	0.01	0.01
UCS Intensity (UCS)	1	21.40*	69.42*	77.62*
CS × UCS	1	2.28	0.04	0.71
CS × R	1	0.06	0.01	1.71
UCS × R	1	0.42	0.31	0.02
CS × R × UCS	1	0.10	4.83	1.46
Within groups (error)	152	(62.62)[a]	(202.04)[a]	(309.63)[a]
Total	159			

* $P < .01$.
[a] Error estimate.

a significant CS × UCS interaction. However, a t test of the difference in CS intensity was significant for the strong UCS groups ($t = 2.90$) but was not significant for the weak UCS groups ($t = 1.16$). From an examination of the curves presented in Fig. 1, it appeared that the strong CS–strong UCS group had approached an asymptotic level at about Trial 50. Since the performance of the other groups had not leveled off at Trial 80, this would have the effect of reducing the CS × UCS interaction since the strong and weak CS groups under the strong UCS would be closer together from Trials 50 to 80 than the strong and weak CS intensity groups under the weak UCS. It was therefore decided to perform a *post hoc* analysis of the CS × UCS interaction over the first 50 trials only. The within-groups variance was used as a basis for evaluating the mean difference (Walker & Lev, 1953), and since the direction of the difference had been predicted a one-sided hypothesis was tested. This test resulted in an interaction value which was not significant ($t = 1.28$, $P = .10$), but which was in the predicted direction, with the difference in performance under CS intensity being greater for the strong than the weak UCS groups.

The groups in which CS intensity was shifted during extinction may be regarded as replications of the CS and UCS conditions during acquisition. The small variance attributable to replication indicates that the random assignment to groups was successful.

A separate analysis of variance was performed on responses falling in the interval from 300 to 565 msec. after CS onset. Spence (1953) has called such longer latency responses "CRs" to differentiate them from the short latency "voluntary" responses. The rationale for the separate analysis of conditioned and "voluntary" responses is evidence from a previous experiment (Taylor, 1954) that these responses may be subject to different laws. In that study, although frequency of "voluntary" responses appeared to be a function of CS intensity, frequency of the longer latency "CRs" did not. Summary of the analysis of "CRs" from the present study is presented in Table 1. Bartlett's test of homogeneity of variance yielded a nonsignificant chi square value. It can be seen from Table 1 that UCS intensity was the important variable, CS intensity having no reliable effect on "CRs."

In Table 1, the results of the analysis of the shorter latency, "voluntary" responses are also presented. Although CS intensity as well as UCS intensity produced significant variance, the assumption of homogeneity of variance was violated. However, recent studies (Lindquist, 1953) have indicated that heterogeneity of variance does not substantially alter the outcome of the analysis of variance test.

Because CS intensity reliably influenced the short latency responses, but not the longer latency responses, an analysis of the form of the responses in the two intervals was made to determine whether the short-interval responses were primarily voluntary in form. The voluntary response shows sharp, smooth closure which is maintained until after the onset of the UCS, whereas the form of the CR is irregular. Table 2 presents the mean number of voluntary and conditioned responses differentiated by form in the two intervals. The responses with voluntary form were fairly evenly distributed between the short and longer latency

TABLE 2

Mean Number of Anticipatory Responses during Acquisition for the Four Conditions According to Form and Latency of Response

Condition		Voluntary			CR		
CS	UCS	150–300 Msec.	300–565 Msec.	150–565 Msec.	150–300 Msec.	300–565 Msec.	150–565 Msec.
Weak	Weak	1.45	0.75	2.20	2.50	22.78	25.28
Weak	Strong	3.25	3.48	6.73	4.60	38.33	42.93
Strong	Weak	2.20	0.58	2.78	4.57	25.23	29.80
Strong	Strong	7.03	6.70	13.73	7.37	38.33	45.70

intervals, with 55% in the short, and 45% in the longer interval. However, for this population, they constituted a much higher proportion of the total responses in the short latency interval (42.2%), than in the longer interval (8.4%).

Separate analyses of variance were done on the responses differentiated by both form and latency. Table 3 summarized the results of these analyses. The voluntary form responses in both intervals were affected by UCS intensity, however only the short latency voluntary responses were influenced by CS intensity. The interaction between CS and UCS intensity was significant at the .05 level for voluntary form responses in both intervals. UCS intensity influenced the CRs in both intervals, but CS intensity affected only the short latency CRs. There was no tendency toward an interaction between CS and UCS intensity for the CRs in either interval.

Extinction of anticipatory response.—The performance curves showing the mean percentage of anticipatory response per block of 6 trials during extinction appear in Fig. 2. Responses occurring in the interval 150 to 2500 msec. after CS onset were considered anticipatory. It can be seen that the curves for the various CS intensity groups under the weak UCS show little trend. The expected decrease in the percentage of anticipatory responses with increased trials is not clearly evident. Only Groups 1 and 3 show a decrement in performance from the first to last extinction

TABLE 3

Analyses of Variance of Total Number of Anticipatory Responses during Acquisition for the Four Conditions According to Form and Latency of Response

Source	*df*	150–300 Msec.		300–565 Msec.		150–565 Msec.	
		Voluntary *F*	CR *F*	Voluntary *F*	CR *F*	Voluntary *F*	CR *F*
CS Intensity (CS)	1	9.91*	8.37*	3.84	0.31	8.74*	2.11
UCS Intensity (UCS)	1	21.24*	8.54*	32.36*	42.90*	36.48*	44.67*
CS × UCS	1	4.43	0.14	4.78	0.31	6.29	0.12
Within groups (error)	156	(20.66)[a]	(28.11)[a]	(24.20)[a]	(191.32)[a]	(65.64)[a]	(251.98)[a]
Total	159						

* $P < .01$.

[a] Error estimate.

trial, with Groups 2 and 4 showing little change in performance over the 30 trials.

In Fig. 2, the performance curves for the strong UCS condition are also presented. Groups 5 and 7 show fairly orderly extinction curves. However, Groups 6 and 8 show little change in performance as a function of increased extinction trials. It can be seen that the strong UCS had the effect of increasing the number of anticipatory responses during extinction, since all groups under the strong UCS are superior in performance to those under the weak UCS.

The results of an analysis of variance performed on the number of anticipatory responses in the interval 150 to 2500 msec. after CS onset during extinction are summarized in Table 4. Analysis of the between-groups variance showed that UCS intensity was the only reliable determinant of response strength during extinction, with CS intensity having no significant effect on performance.

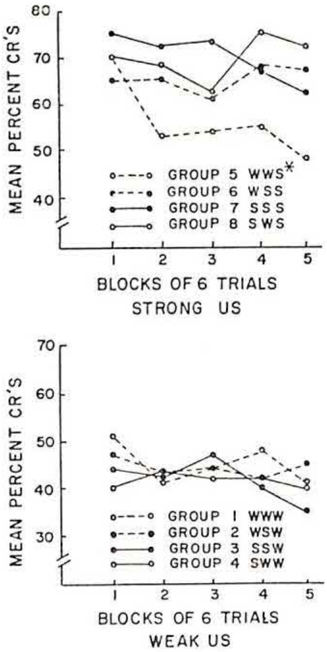


FIG. 2. Mean percentage of responses in the interval 150 to 2500 msec. during extinction per block of 6 trials as a function of CS and UCS intensity. In the legend, the first letter stands for CS intensity during acquisition (W = weak, S = strong). The second letter stands for CS intensity during extinction. The third letter stands for UCS intensity.

TABLE 4

ANALYSIS OF VARIANCE OF TOTAL NUMBER OF ANTICIPATORY RESPONSES IN THE INTERVAL 150 TO 2500 MSEC. AFTER CS ONSET DURING EXTINCTION FOR THE EIGHT GROUPS

Source	*df*	*MS*	*F*
CS Intensity, Acquisition (CS-A)	1	37.06	
CS Intensity, Extinction (CS-E)	1	18.91	
UCS Intensity (UCS)	1	1735.81	16.70*
CS-A × UCS	1	127.81	1.23
CS-A × CS-E	1	12.66	
UCS × CS-E	1	23.26	
CS-A × CS-E × UCS	1	24.78	
Within groups (error)	152	103.93	
Total	159		

* $P < .01$.

A simple analysis of variance was done using the same criterion for an anticipatory response as that used during acquisition, i.e., any response which occurred in the interval 150 to 565 msec. after CS onset. Since the between-groups variance was not significant, no further analysis of these data was undertaken.

DISCUSSION

Acquisition.—The acquisition results of this experiment have provided clear-cut evidence of a positive effect of CS intensity on performance and are therefore consistent with a part of Hull's postulate of stimulus intensity dynamism. In eyelid conditioning, CS intensity has not previously been shown to influence response strength. It was pointed out earlier that the negative findings of these studies may have been related to the level of the UCS employed. In this experiment, the main hypothesis that CS intensity would have a greater effect on performance under strong UCS intensity than under weak UCS intensity was not clearly verified since the CS × UCS interaction was not significant. However, the fact that the difference between strong and weak CS intensity was significant under the strong UCS but was not reliable under the weak UCS is at least consistent with the hypothesis. Further evidence that UCS intensity may be one of the conditions which influences the effectiveness of CS intensity on behavior comes from the results of the analysis based on the first 50 trials. Although significant at only the .10 level, this result at least suggests that the interaction between CS and UCS intensity may be a real effect.

It may be seen from Table 1 that response strength is more determined by UCS intensity than CS intensity. The finding that a strong level of UCS reliably facilitated performance is consistent with previous studies (Passey, 1948; Spence, 1953; Spence & Taylor, 1951) which have shown UCS intensity to influence acquisition of the conditioned eyelid response. Since the effect of CS intensity on performance is relatively small, a greater number of *S*s is required in order to demonstrate it than in the case of UCS intensity. This may be related to the fact that the results of studies on UCS intensity in eyelid conditioning have been more consistently positive than those on CS intensity.

The analysis of responses by form showed that the tendency toward an interaction between CS and UCS intensity appeared to be due almost entirely to the voluntary form responses, and did not seem to depend on latency. The analysis also showed that for CS intensity, the important factor is the latency of the response rather than the form. CS intensity influenced both voluntary form and CRs in the short latency, but not the longer latency interval. The finding that CS intensity affected the short latency responses but not the longer latency responses is in agreement with the results of a previous study of the eyelid response (Taylor, 1954). Further evidence that the short and long latency responses may be subject to different laws is the finding that UCS intensity had a reliably greater effect on the longer latency responses than the short latency responses. A 2 × 2 × 2 factorial analysis with UCS, CS, and latency as variables yielded significant values for the three main variables and a significant UCS × Latency interaction ($P < .01$). The short latency responses did, however, behave like the longer latency responses in showing a significant increment during acquisition under the strong CS-UCS condition as indicated by an analysis of trend (Lindquist, 1947). It is difficult to evaluate these results since there has been no systematic study of the variables which influence the differentiated form and latency responses. The results of the present experiment, however, suggest that such an investigation may prove worthwhile.

Extinction.—Since the extinction results showed that CS intensity during acquisition had no differential effect on performance, this indicated that there was no reliable effect on learning or habit strength. This interpretation is possible because the factorial design used ruled out differential effects of CS intensity on response strength. The present experiment, therefore, offers no support for the part of Hull's stimulus intensity dynamism postulate (V_1) which specifies that CS intensity affects habit strength.

Although the acquisition results supported the stimulus dynamism prediction that CS intensity would affect performance or reaction potential, the extinction results did not, since the variance associated with CS intensity during extinction was not significant. It is difficult to explain why CS intensity should affect acquisition but not extinction of a CR. Kessen (1953), who reported a similar finding in an instrumental conditioning situation, suggested that the decreased drive resulting from omission of the UCS during extinction could have been a factor. In this experiment, however, such an explanation is inadequate. The UCS was presented on all extinction trials and that the drive level was maintained can be inferred from the significant difference in performance under strong and weak UCS during extinction.

The gradual, continuous decrease in performance which McAllister (1953) obtained was shown only by the two groups under the strong UCS in which CS intensity remained the same from acquisition to extinction. Since the two groups in which CS intensity was shifted failed to show the expected decrement it may be that shifting CS intensity reduced extinction. Another possibility is that the gradual response decrement did not appear for the shifted groups because of a depressed response level at the outset of extinction associated with stimulus generalization. On the first extinction trial the shifted groups were reliably inferior in performance to the nonshifted groups.

The lack of a general decreasing trend in performance for the groups under the weak UCS is understandable in view of McAllister's finding that for the 2500-msec. interval employed there is a random response rate of about 40%. Since in the present study the groups under the weak UCS began extinction with low performance levels of 40–50%, there was little opportunity for extinction in these groups. Under the strong UCS condition the groups showed a response level considerably higher at the end of extinction than the 40% random rate obtained in the McAllister study. There was a methodological difference between the two studies since McAllister used a partial reinforcement procedure with 20% test trials during acquisition and extinction where the UCS was omitted, whereas in the present study the UCS was presented on all trials. It seems unlikely that this could have accounted for the higher final response level in the strong UCS groups in the present study since partial reinforcement has consistently increased resistance to extinction (Jenkins & Stanley, 1950). It appears more reasonable that the higher performance was associated with increased drive resulting from the stronger UCS employed in the present study.

Although the acquisition results of this experiment were consistent with the stimulus intensity dynamism postulate, the extinction results offered little support for it. To the extent that the finding of a CS intensity effect on performance during acquisition could not be duplicated during extinction, it would appear that the extinction data may not have provided a fair basis for evaluating that part of the hypothesis which concerns habit strength.

SUMMARY

The present experiment investigated the effect of CS intensity on acquisition and extinction of the conditioned eyelid response under different levels of UCS.

Eight groups of 20 male *S*s each, were given 80 acquisition trials under four different stimulus conditions: weak CS–weak UCS; strong CS–weak UCS; weak CS–strong UCS; strong CS–strong UCS, with two groups under each condition. Each group was then administered 30 extinction trials with the CS-UCS interval shifted from 500 to 2500 msec. All groups received the same UCS intensity as during acquisition with four groups receiving the same CS intensity as in acquisition and the other four groups being shifted from either weak to strong, or strong to weak CS intensity.

The results showed that CS intensity reliably influenced response strength during acquisition. CS intensity affected the short latency, voluntary form, and conditioned responses, but not the longer latency responses. The latter, however, were more influenced by UCS intensity than the former. There was some evidence that CS intensity had a more pronounced effect on performance under a strong UCS than under a weak UCS. This effect was restricted to the voluntary form responses. The extinction results showed no evidence of an effect of CS intensity on either learning or performance. However, performance during both acquisition and extinction was a function of UCS intensity.

In general, the results of this experiment supported the part of Hull's stimulus dynamism principle which postulates an effect of CS intensity on response strength. There was, however, no evidence of an effect of CS intensity on learning as distinct from performance.

REFERENCES

BARNES, G. W. Conditioned stimulus intensity and temporal factors in spaced-trial classical conditioning. *J. exp. Psychol.*, 1956, **51**, 192–198.

BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.

CARTER, L. F. Intensity of conditioned stimulus and rate of conditioning. *J. exp. Psychol.*, 1941, **28**, 481–490.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.

HOVLAND, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–290.

HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

KESSEN, W. Response strength and conditioned stimulus intensity. *J. exp. Psychol.*, 1953, **45**, 82–86.

LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953. Pp. 78–90.

MCALLISTER, W. R. The effect on eyelid conditioning of shifting the CS-US interval *J. exp. Psychol.*, 1953, **45**, 423–428.

PASSEY, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.

PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

RAZRAN, G. The dominance-contiguity theory of the acquisition of classical conditioning. *Psychol. Bull.*, 1957, **54**, 1–46.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

SPENCE, K. W., & TAYLOR, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TAYLOR, E. Generalization of the conditioned eyelid response to an auditory stimulus varying in intensity. Unpublished doctoral dissertation, Univer. of Iowa, 1954.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953. Pp. 158–160.

(Received June 3, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

TRACKING PERFORMANCE ON A SEQUENCE OF STEP FUNCTIONS WHICH APPROACHES A CONTINUOUS FUNCTION AS A LIMIT[1]

DAVID McCONNELL AND MAYNARD W. SHELLY[2]

Ohio State University

A number of previous studies (Battig, Gregg, Nagel, Small, & Brogden, 1954; Battig, Voss, & Brogden, 1955; Craig, 1949; Craik, 1948, Ellson, Harris, & Craig, 1949; Ellson & Hill, 1948; Ellson & Wheeler, 1949; Hartman & Fitts, 1955; Searle & Taylor, 1948) have undertaken to relate stimulus amplitude and response amplitude in tracking tasks. In particular, the Battig studies investigated the effect of an increase in frequence of target intermittency on tracking error and found in the later study (Battig, Voss, & Brogden, 1955) a linear relation between this increase and proficiency of tracking.

The present study investigates pursuit tracking performance as the discrete movements of a target on a cathode ray oscilloscope (CRO) increase in number and decrease in amplitude. The discrete target movements approach the limiting continuous movement in a manner analogous to a discrete function $D_n(x)$ uniformly converging to some continuous function $C(x)$, where n represents the number of steps, and x the displacement of the target with respect to the center of the display.

Two important differences between the Battig et al. studies and our own, which affect the extent to which results may be compared, are: (*a*) the use of a step function rather than an intermittent function does not entail possible influences on the response by factors such as brightness or stimulus perseveration associated with intermittent stimuli, and (*b*) the present experiment systematically varies the frequency of the underlying target and also includes a random target condition in order to provide more information about the generality of the relation between stimulus discontinuity and tracking error. In comparing the Battig et al. *limiting* target, which was dependent for perceptual continuity on a rate of intermittency in excess of the fusion rate, with our continuous target, there is little or no reason to believe that these two modes of presentation would, in and of themselves, affect tracking error differently.

METHOD

Apparatus.—The apparatus used in this study is depicted in Fig. 1. The continuous function generator consisted either of a Hewlett-Packard variable-frequency oscillator or of a gas-tube random signal generator with mean frequency of 10 cpm. Under *continuous* conditions the output of the function generator was fed directly to the mixer, and from thence to the display and scoring circuits, while under *discrete* conditions it was introduced into the quantizer, which broke the signal up into particular amplitudes and distributions of steps. Components including

---

[1] This research was carried out in the Laboratory of Aviation Psychology and was supported in part by the United States Air Force under Contract AF 41(657)-70 with the Ohio State University Research Foundation, monitored by the Aero Space Medical Laboratory, Wright Air Development Center. Permission is granted for reproduction, publication, use, and disposal in whole and in part by or for the United States Government. The aid of Paul M. Fitts, now at the University of Michigan, is gratefully acknowledged by the authors.

[2] Now with the Office of Naval Research, Chicago, Illinois.

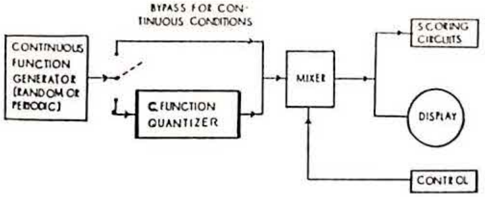

BYPASS FOR CONTINUOUS CONDITIONS
CONTINUOUS FUNCTION GENERATOR (RANDOM OR PERIODIC)
C FUNCTION QUANTIZER
MIXER
SCORING CIRCUITS
DISPLAY
CONTROL

FIG. 1. Flow diagram of the apparatus.

and to the right of the mixer in Fig. 1 are those of Electronic Pursuit Apparatus (EPA) III, an earlier version of which has been described in detail elsewhere (Warren, Fontaine, & Clark, 1952). In the present application EPA was used to present the target and cursor signals on a 5-in. cathode ray oscilloscope (CRO) and to generate various measures of the discrepancy (error) between the two signals. The target on the CRO appeared as a line 18 mm. long and .5 mm. wide, and moved horizontally across the face of the scope. The top of the cursor line was displaced 10 mm. below the top of the target line and was of similar dimensions. Together, when in alignment, they formed a line 28 mm. long. The surface of the CRO was about 28 in. from *S*, 10° below eye level, and perpendicular to the line of sight. The target and cursor were painted on the display by a 200-cps time-sharing switch. Except for large discrete jumps, this produced no stroboscopic effects.

The room in which *S* served was dimly illuminated. Specular reflection from the CRO was eliminated by means of Polaroid filters. The *S* sat in a chair which could be adjusted for height. The elbow of his right arm rested on the pivot point of a horizontal lever and his fingers grasped a knob at the end of the lever. Attached to the control was a potentiometer. The voltage across this potentiometer determined the position of the cursor on the CRO. A maximum movement of the control 40° to the right or left corresponded to a movement of the cursor on the CRO of 1.5 in. to the right or left, respectively. The arm control was of light construction, and was not spring-centered.

Included in the scoring circuits was an amplitude analyzer (Erdman, Fontaine, & McConnell, 1959) designed to record the amplitude distribution of errors generated by *S* when attempting to track the target. This device breaks a continuous error function (a voltage signal) down into a discrete amplitude distribution by measuring the error voltage at .1-sec. intervals and recording a count, proportional to the amplitude category of this voltage, on a cumulating counter corresponding to that particular amplitude.

Normally, this apparatus recorded errors to the left and to the right of the target in appropriate counters. However, it was possible to throw a switch so that the apparatus would record errors which were dependent upon the direction of the target for all signals which did not change direction between points of maximum excursion. When

TABLE 1

DESIGN OF THE EXPERIMENT

Number of Steps into Which Range is Divided	Distribution of Discrete Step Inputs	Frequency of Signal			
		Triangular			Random
		10 cpm	20 cpm	30 cpm	10 cpm
2	Equal	T-2-0-10	T-2-0-20	T-2-0-30	R-2-0-10
4	1	T-4-1-10	T-4-1-20	T-4-1-30	R-4-1-10
	2	T-4-2-10	T-4-2-20	T-4-2-30	R-4-2-10
	3	T-4-3-10	T-4-3-20	T-4-3-30	R-4-3-10
8	1	T-8-1-10	T-8-1-20	T-8-1-30	R-8-1-10
	2	T-8-2-10	T-8-2-20	T-8-2-30	R-8-2-10
	3	T-8-3-10	T-8-3-20	T-8-3-30	R-8-3-10
12	Equal	T-12-0-10	T-12-0-20	T-12-0-30	R-12-0-10
Continuous	—	T-C-0-10	T-C-0-20	T-C-0-30	R-C-0-10

this switch was in the lead–lag position, it recorded counts in one band of counters when S was ahead of (led) the target, and in another bank of counters when S was behind (lagged) the target. On odd-numbered trials this switch was in the lead–lag position, and on even trials in the left–right position.

Subjects.—Five naive *S*s were used. All were male, right-handed students who received hourly wages.

Design.—The experiment consisted of 36 conditions. All *S*s served six times in each condition. The 36 conditions are presented in Table 1.

As indicated, two types of underlying signals were used—a random signal and a triangular signal. The triangular signal was used because the average increment of its corresponding step input is linearly related to the theoretical error between continuous and discrete signals. The random signal had a normal distribution of amplitudes and its power spectral density was peaked at 10 cpm. The frequencies of the triangular signal were 10, 20, and 30 cpm. The range of both functions was divided into 2, 4, 8, and 12 discrete steps, or was continuous. For 4- and 8-step functions, one of three distributions of the steps (relative to the mean increment size) was used. The sets of numbers and letters in the cells of the design (Table 1) give the code by which the individual conditions were designated. In each cell, the first position refers to the wave form ("T" for triangular, and "R" for random). The second position refers to the number of steps into which the signal is quantized. When a continuous function is used, a "C" stands in the second place. The third position refers to the particular distribution of increments used in quantizing the function. If all the increments were equal, an "O" is in this position. The last position refers to the frequency of the underlying function in cycles per minute. Thus, for example, T-8-2-20 refers to the triangular eight-step, second distribution, 20-cpm function. In the bottom row of Table 1, the distribution index (0) can be ignored since the functions are continuous.

In the distribution of the amplitudes of the step inputs for all numbers of discrete inputs used in this study, the same maximum excursion was maintained. The amplitudes of the step inputs were all equal for the 2-step and 12-step conditions. For the 4- and 8-step conditions, however, a variety of distributions was used.

Procedure.—When S first reported to the laboratory he was seated on S's stand facing the CRO, and given the following instructions:

> On the scope in front of you are two signals. One is moving. We will call it the target. (The target was T-C-0-10, Table 1.) The other is stationary. We will call it the cursor. Now take the handle in your right hand. Make your arm comfortable. Now note that you can move the cursor by moving the arm control. (E moved S's arm back and forth to demonstrate, making a rough effort to track the target. After a few seconds, he removed S's hand from the control.) You will see several different kinds of targets: some will be chopped up and some will move smoothly; some will move faster than others. Your job is to keep the horizontal distance between the target and cursor at a minimum at all times. Each period will be just over a minute long, and at the start of each period. I will say 'ready,' from the next room, and if you say 'yes,' we will begin. Any questions?

Each trial was of 65-sec. duration, the first 5 sec. being an unscored warm-up period. Between trials S rested 2 min., except that he was permitted two 10-min. breaks in a day's run. Each S served for an average of 27 trials each day, or a little less than 2 hr., including rest periods. This continued for eight days. During the experiment each S went through each condition six times. The order each time was random with the obvious restriction that conditions were not repeated until S had served in all 36 conditions an equal number of times.

Measures.—The theoretical error between the periodic step function and the continuous function was derived as follows: The distance between the step function and the continuous function was squared and its definite integral with respect to time computed. After averaging, the square root was then taken. The result is symbolized by $\mathrm{E} = (\mu_2)^{\frac{1}{2}}$ where μ_2 is the usual symbol for the second moment about the origin which, in this case, was taken as the continuous signal.

On every trial two voltage readings were taken. The first represented the voltage difference between the cursor and the target (T) squared and integrated over the last 60 sec. of each trial. The second reading represented the squared and integrated voltage difference between the cursor and the underlying continuous function (C). When S was tracking the continuous function the two voltage readings should have been the same, but because of limitations of the scoring circuits it was not possible to maintain this condition for all input voltages. Consequently, a curve relating the two circuits was plotted as a function of input voltage, and from it the two voltmeter readings were adjusted so that the scale factors were the same for both scores. This curve was linear over most of the range.

The first score is represented by E_T and the second one by E_C. Symbolically, if the difference between the signal and cursor at Time t is e_T or e_C, the condition (T or C) and time interval over which the score is integrated by I_T or I_C, then:

$$E_T = \left[\frac{\int_I e_T^2(t)dt}{I_T}\right]^{\frac{1}{2}} = (\mu_2, C)^{\frac{1}{2}} \qquad [1]$$

where t stands for time. Similarly,

$$E_C = \left[\frac{\int_I e_C^2(t)dt}{I_C}\right]^{\frac{1}{2}} = (\mu_2, C)^{\frac{1}{2}} \qquad [2]$$

Note that neither E_T nor E_C is strictly rms error *unless* *S*'s error is centered about the target.

The two voltage readings were then converted to inches measured on the face of the scope. This was possible through a knowledge of the constants of integration for the meters. Such a conversion of scores was linear. As a further check on the accuracy of the conversion, the converted scores were compared with similarly defined scores computed on the basis of the amplitude analyzer data. The average absolute difference between the converted meter scores and the scores based on the amplitude analyzer for a sample of scores was 4.57%. The average error of comparison for this same sample was .39%.

Amplitude analyzer data were collected on every trial. For the random function this entailed the recording of errors to the left or right. For the triangular function, on even-numbered presentations of the conditions, left–right errors were recorded, and on odd-numbered presentations of the conditions, lead–lag data were collected. Graphic records were taken periodically of target, cursor, and error voltages in an effort to determine at what number of steps (if any) *S* changed from one mode of responding (i.e., continuous or discrete) to the other.

RESULTS

Although measures were taken throughout the six cycles, results reported are based only on the last five complete cycles when *S* was highly practiced. Because of this, the error term was small enough to permit the use of parametric procedures when needed. Since, however, Wilcoxon's paired replicates test was appropriate to many situations in which it was desired to test differences, this nonparametric test was the most extensively used. As too many of the scores for R-2-0-10 (the 2-step condition with a random underlying signal) reached the maximum of the scoring devices, the scores obtained for that condition seriously underestimate the true scores. This was not true of any other conditions. Therefore, R-2-0-10 was not included in the analyses. Whenever a result is referred to as significant, the .05 level may be assumed.

Figure 2 shows the average error (E_T) with respect to target as a function of the target step size (θ). Results are very similar for the error (E_C) with respect to the underlying continuous signal. In the figure the three distributions of step input amplitudes have been averaged for both 4- and 8-step conditions. Because of the concern in this study with the approach to a limit, it is relevant to raise the question as to whether or not the findings are specific to the type of distributions used in these conditions. The conclusion seems to be negative on several grounds. First, it might be suspected that an even distribution, because of its closer approximation (smaller E) to the continuous triangular function, would lead to smaller error scores. Approximation to the underlying continuous function did not, however, linearly order the resulting error on the different distributions. This held for both E_T and E_C. Furthermore, using Wilcoxon's paired replicates test and considering only those pairs which have the same underlying continuous distribution, of the 24 pairs which were tested only two showed significant differences. These were both for 4-step conditions. Furthermore, with the data upon which Fig. 2 is based, when linear functions were fitted to the data points without first averaging the values obtained for different distributions, no significant differences were obtained between the obtained slope or intercept estimates when Wilcoxon's paired replicates test was used. It will also be recalled that the data underwent a number of transformations before being fitted with linear functions. Only one of these for one meter, however, could have altered the values so that they could be better described by a linear function. The unaltered scores obtained from this meter (for E_C) were plotted and fitted with linear functions. As was found with the transformed scores, the data points could be described by linear functions with the error substantially between 2% and 3%. Thus there are a number of strong reasons for believing that the functions obtained are estimates of general relation, and are *not* dependent upon the precise distribution of steps in the discrete function.

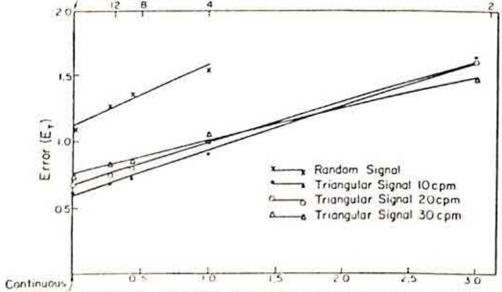


FIG. 2. Average error (E_T) with respect to target as a function of step input amplitude (θ).

In Fig. 2 all of the sets of points corresponding to a single underlying continuous signal have been fitted with linear functions. A least-squares technique was not used for two reasons. It was felt that some of the extreme deviations could as easily be the result of apparatus malfunctioning as a result of the behavior of *S*s. Data immediately preceding any apparatus failure were not included, but this precaution was undoubtedly not foolproof. Second, it was felt that since curves were being fitted to the data of individual *S*s before being averaged, the least-squares method would add unduly to the variance and affect the final estimates somewhat more. Therefore, the method of successive differences was used.

Two points emerge from inspection of Fig. 2. The first is that both the slope and intercept of the function describing performance on the random signal conditions are greater than these constants for the triangular conditions. This difference was substantiated by Wilcoxon's paired replicates test, using trials as the basis for matching scores. In both cases, that triangular function which led to performance closest to that on the random conditions was used for comparison.

The second point is that the slope of the function fitting the 10-cpm triangular signal data is greater than the slope of the function fitting the 30-cpm data. This difference was tested in a similar way and found to be significant. Since these two functions intersect at 1.75-in. step amplitude, it follows that opposite relations between frequency of the underlying continuous signal and error obtain to the left and to the right of the intersection; that is, for lower amplitudes, increasing signal frequency from 10-cpm to 30-cpm increased error, while for higher amplitudes the frequency increase resulted in lower error.

The distribution of errors.—The distributions of errors obtained from the amplitude analyzer indicated two things: (*a*) that for all conditions *S*s were on the average to the left of the target, and (*b*) that for all triangular conditions *S*s were leading the target with respect to its prevailing direction (Fig. 3). Lead and lag information was not determined for the random functions because of an instrument limitation. Significance levels indicated on Fig. 3 were determined by conducting sign tests on the equality of distribution frequencies in certain critical zones of counters. A dash adjacent to a function indicates that the effect of increasing the number of steps was nonsignificant at the .05 level. The two significance levels indicated by vertical arrows are associated with similar sign tests to detect distributional differences attributable to frequency for these functions. The conversion from the distributional frequencies to inches on the CRO was accomplished by reducing the apparent mode of a particular distribution to a counter number, then referring to a conversion table computed from previous calibration data. When distributions were multimodal, they were smoothed by fitting a polynomial, and the apparent mode was then taken as the counter number nearest to the curve maximum.

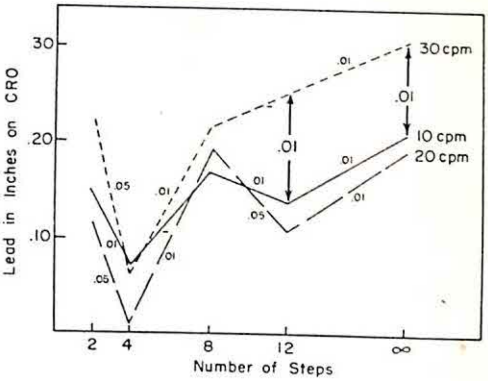


FIG. 3. Lead on the CRO display as a function of the number of steps. The decimals adjacent to the functions indicate the outcome of sign tests of the distributional differences in selected groups of counters.

Figure 3 shows that the mean amount of lead increases with the number of steps, but the levels of significance between adjacent increments make it difficult to ascribe a single uniform function to this relationship. In fact, it would appear that for relatively low step numbers, frequency does not order the amount of lead, while for relatively high step numbers (including the continuous case) there is a significant effect of frequency on lead. On the other hand, the number of steps appears to exert a significant effect on the amount of lead, almost regardless of frequency. In particular, the 4-step function appears to be associated with the smallest lead.

A remarkably consistent relationship appeared in comparison of error distributions for 4-, 8-, and 12-step random and triangular step functions. For the 2-step and continuous functions, consistent differences between the two types of functions did not appear. For the others the random function clearly shifted errors to the left of the target—significantly further to the left than for the triangular function.

DISCUSSION

The findings indicate that over the sequence of experimental conditions the relationship between average amplitude of the step input and the particular error measure selected was linear. Put another way, as the sequence of step inputs more closely approximated the underlying continuous signal, the error measure approached the limiting error. The limiting error was, of course, well above zero—that is, *S*s responded to neither discrete nor continuous targets with perfect accuracy.

The effect of change in frequency upon *S*'s ability to track the periodic function was merely to change the slope of the linear function. Moreover, for two very diverse underlying signals (triangular and random) the form of the error function remained the same; i.e., *the form of the approach to the behavioral limit within these conditions was independent of the character of the limiting signal.* Thus, for the measure used, the extent to which *S* could track a discontinuous signal could be predicted simply by a knowledge of his performance on the corresponding continuous signal and the slope constant of the linear function. Moreover, this common property to both periodic and random signals emerges despite the unsurprising fact that use of the random input clearly degrades performance with respect to the periodic input.

There is evidence, e.g., from graphic records, that even at eight steps with the high-frequency periodic signal *S*s make distinct responses. Yet not until the number of steps is reduced to four, does the particular distribution of a fixed number of steps across the tracking range appear to have any differential influence on error. This means that *S* can compensate without degradation of his response, for variation in the positions of discrete stimuli along the underlying path, up to some limit (here, apparently, four steps). What the determinants are of this "limit," beyond which *S* loses ability to "smooth" the function equally well for all distributions, is a question open to speculation. Taylor (1948) has stated that he believes *S*s develop sets for responding to the average size of error (average amplitude) and that this influences all of his responses in the direction of the average. A number of later investigations (Craig, 1949; Ellson & Hill, 1948; Ellson & Wheeler, 1949) have confirmed this, one of them (Ellson & Wheeler, 1949) showing that the effect is due to the range and not to the absolute amplitudes of the step inputs. These findings would seem to indicate that *S*s, by overshooting small amplitude inputs and undershooting large amplitude inputs, would increase their error. The Taylor theory appears to be confirmed in the case of the 4-step function. The larger errors were found, in our experiment, in both cases for the more irregular distribution—where the signal appeared farther from the average signal, and also from the corresponding position of the underlying signal.

In an effort to correlate the lead–lag errors of Fig. 3 with the integrated error of Fig. 2, we offer the postulate that for small to moderate ranges of increment size, *S* minimizes his errors as target difficulty increases by increasing his lead on the target. Under this general hypothesis, the following points arise:

1. The fact that at two steps error is greater than at four, while lead is also greater, can be attributed to the fact that between the two extremes of the 2-step function there were almost 100° of arm movement, which meant that to keep zero error, *S* would have had to stay on the target at one extreme, then move instantaneously to the other extreme. Since this is impossible *S* is confronted with two alternatives—either to stay with the target until it moves, thereby accepting a lag error with every movement, or to try to anticipate by a fraction of a second when the target will move. That *S* chooses the latter course *in the absence of instructions* as to lead–lag behavior is clear from the data. While his error is maximal under this extreme condition it may very well be that within the condition, the anticipatory or *lead* response is still the most efficient—i.e., produces less error than any other mode of response. An obvious experimental check on this inference is to instruct two groups of *S*s differently: the first to anticipate when the target will move by taking a lead on it just before it moves, as *S*s appear to have done in this study, and the second to stay on the target at all costs until it moves.

2. The decrease in lead from eight steps to twelve for 10- and 20-cycle targets, corresponding to the decrease in error, is together with graphic records interpreted by

*E*s as demonstrating that the mode of response changes from discrete to continuous across the step change 8 to 12. Because of the deliberate omission of instructions relevant to continuous versus discontinuous responding, it is concluded that the optimal mode at eight steps for the two lower frequencies was still discrete, while it was continuous for the 30-cycle target. At 12 steps, the optimal mode appears to have been continuous with a lessening of lead. At this number of steps, no discrete element could be detected in the graphic record for any *S*, for any frequency.

3. The discrepancy between the 30-cpm function and the 10- and 20-cpm functions at 12-step and continuous values of the argument (Fig. 3) corresponds to the finding already discussed (Fig. 2) that 30-cpm error appears to be slightly greater for these arguments. Since the targets are more difficult to track, *S* evidently increases his lead, and, as above, thereby reduces his error to a minimum for that particular target.

It is obvious, therefore, that explicit confirmation of the hypothesis that *S* improves his performance by leading the target must be contingent upon use of targets of equal difficulty, while constraints are experimentally imposed upon lead and lag behavior.

One further key to the interpretation of the lead results is the singular relationship indicated above that as the target increases in difficulty, *S* apparently stays increasingly to its left. Our hypothesis here is that *S* puts most of his lead into the cursor during flexion. (Right-handed *S*s flex to the left.) This notion is clearly complementary to the theory that leading the target improves performance, since we know that flexion responses are, in general, anatomically easier to control. Since we were unable to make comparisons of lead–lag and left–right distributions on the same trial, the experimental accreditation of this hypothesis is not entirely established.

SUMMARY

The relation between error in tracking a continuous function and error in tracking a discrete function which approaches the continuous function in the limit was examined. Random and periodic targets were presented to five *S*s, and response measures included root-mean-square error between target and cursor, as well as between cursor and the limiting continuous function. Also measured were average lead–lag error of the cursor with respect to the target, and average left–right error of the cursor with respect to the target. Results indicated that as the number of steps is increased, the error measure chosen on the discrete targets linearly approached that associated with the limiting continuous target. This linearity was virtually independent of frequency or whether the limiting function was periodic or random. The *S*s on the average led the target under all conditions, and were also to the left of the target on the average. The relation of these results to the previous finding was interpreted as indicating that *S* minimizes his error by leading the target, and that this is especially pronounced during flexion of the control arm.

REFERENCES

BATTIG, W. F., GREGG, L. W., NAGEL, E. H., SMALL, A. M., JR., & BROGDEN, W. J. Tracking and frequency of target intermittence. *J. exp. Psychol.*, 1954, **47**, 309–314.

BATTIG, W. F., VOSS, J. F., & BROGDEN, W. J. Effect of frequency of target intermittence upon tracking. *J. exp. Psychol.*, 1955, **49**, 244–248.

CRAIG, D. R. Effect of amplitude range on duration of responses to step function displacements. *USAF Tech. Rep.*, 1949, No. 5913.

CRAIK, K. J. W. Theory of the human operator in control systems: II. Man as an element in a control system. *Brit. J. Psychol.*, 1948, **38**, 142–148.

ELLSON, D. G., HARRIS, H., & CRAIG, D. R. The interaction of responses to step function stimuli: II. Equal opposed steps of varying amplitude. *USAF Memo. Rep.*, 1948, No. MCREXD-694-2P.

ELLSON, D. G., & HILL, H. The interaction of responses to step function stimuli: I. Opposed steps of constant amplitude. *USAF Memo. Rep.*, Nov. 1948, No. MCREXD 604-2P.

ELLSON, D. G., & WHEELER, L., JR. The range effect. *USAF Tech. Rep.*, 1949, No. 5813.

ERDMAN, A. C., FONTAINE, A. B., & MCCONNELL, D. G. The amplitude analyzer. *USAF WADC Tech. Rep.*, July 1959, No. 59–45.

HARTMAN, B. O., & FITTS, P. M. Relation of stimulus and response amplitude to tracking performance. *J. exp. Psychol.*, 1955, **49**, 82–92.

SEARLE, L. V., & TAYLOR, F. V. Studies of tracking behavior: I. Rate and time characteristics of single corrective movements. *J. exp. Psychol.*, 1948, **38**, 615–631.

TAYLOR, F. V. Certain characteristics of the human servo. AIEE Misc. Paper 48-235, 1948.

WARREN, C. E., FONTAINE, A. B., & CLARK, J. R. A two-dimensional electronic pursuit apparatus. *USAF Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52–26.

(Received June 3, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

MULTIDIMENSIONAL MAPPING OF MUNSELL COLORS VARYING IN HUE AND CHROMA[1]

TAROW INDOW AND TSUKIKO UCHIZONO

Keio University, Tokyo

The purpose of this study was to make an experimental inquiry, by the use of the multidimensional scaling, into the possibility that differences perceived among colors would be adequately represented by distances in a real Euclidean space. Here, by color differences, not only ones slightly above the difference limen but also supraliminal ones of considerable degree are to be understood.

As is well known, in the Munsell system, a surface color is specified by a point in a solid in which cylindrical coordinates represent, by definition, the principal attributes of colors. Any horizontal section through the solid is defined as a plane of constant lightness (Munsell value), while any vertical plane originating at the achromatic axis represents a plane of constant hue, and a cylindrical section concentric with the axis constitutes a surface of constant saturation (Munsell chroma). The scaling units for the three attributes are defined so that equal distances on each of the hue, chroma, and value scales would represent equal observed visual intervals of that attribute. Then, one question will arise. Suppose that two colors are given which are respectively 5R 6/2 and 5R 4/8 in the Munsell notation. It is obvious that these colors differ by 2 units in value and by 6 units in chroma. It is not clear, however, how to represent the total impression of difference between these colors. Is it given by the magnitude $\sqrt{(w_1 2)^2 + (w_2 6)^2}$? In the reports of the OSA subcommittee (Newhall, 1940; Newhall, Nickerson, & Judd, 1943) which examined the spacing of Munsell colors in detail, it was stated that no attempt was made to evolve a single scaling unit for the entire solid, or to equate the scaling unit for hue, value, and chroma. Hence, the relative size of these units, e.g., w_1 and w_2, is not known, although several tentative definitions of the relative size have been made (Godlove, 1951). Furthermore, it is not self-evident that the distance corresponding to the difference over more than one attribute is Euclidean in nature and also that value and chroma, for example, are to be represented by axes orthogonal to each other.

The Euclidean character of the Munsell solid is tacitly assumed in various formulae to evaluate color differences quantatively (Judd, 1952). These formulae are regarded, in general, as applicable only to differences among more or less similar colors. That means only the differences slightly above the limen have been taken into consideration when experiments were undertaken to test the adequacy of the Euclidean model in the Munsell system. The same is true in experiments for developing a uniform tridimensional color scale. It has been of primary concern in these experiments to examine whether colors are arranged in a tridimensional array so that each color differs from each of its nearest neighbors by

---

[1] This research was supported by a grant from the Ministry of Education, Japan. The authors are indebted to M. Matthewson for her helpful advice concerning English.

perceptually equal amounts. In short, uniformity of the scale *in the small* only has been the subject of the experiments, and little attention has been paid to uniformity of the scale *in the large.*

Torgerson (1952, 1958) developed the method of multidimensional scaling and applied it to a set of nine Munsell colors all of the same red hue (5R) but differing from each other in value and chroma. For obtaining comparative distances among these colors the complete method of triads was used in his experiment. The experiment was repeated by Messick (1954, 1956) with different procedures for obtaining distances among colors again varying in value and chroma but kept roughly at 5R in hue. The results of the multidimensional mapping were much the same in both experiments and the configuration of colors agreed closely with the Munsell system. Richardson (1938) obtained essentially the same results, though the procedures of multidimensional mapping were simpler than those of Torgerson and Messick. It is to be noted that color differences *in the large* play an important role in the multidimensional scaling. Torgerson and also Messick pooled the responses of a number of *S*s and obtained a common set of distances among colors to which the multidimensional scaling was applied. The close agreement of the final results with the Munsell system would, therefore, hold only in the responses of *S*s as a whole but not necessarily in the responses of each *S*. It is open to question especially because both methods of scaling the distances presuppose heterogeneity within *S*s. That is to say, the existence of individual differences as to what *S*s perceive between a pair of colors is requisite for the procedures. Such individual differences, however, might be qualitative as well as quantitative in nature. Employing the same nine colors as used by Torgerson, Indow and Shiose (1956) constructed the individual multidimensional configuration of the colors for each of three *S*s. The configurations obtained were in good agreement with each other and also with the Munsell system. In this experiment, the nine colors were of the same red (5R) but varied in value and chroma, from 3 to 7 and 4 to 12 respectively. Variation in hue, from 6R to 6YR, was added to that in value and chroma in the experiment carried out by Shiose (1958). Tridimensional configuration of the colors was successfully constructed in agreement with the Munsell system though the range of hue was rather restricted. In the present experiment, 21 colors were used which varied in hue and chroma over the widest possible range. The lightness was kept constant at 5 in Munsell value. As will be discussed later, among the basic postulates of the Munsell system the most questionable one may be that hue is represented by a radius vector and chroma by a concentric circle in a Euclidean plane.

METHOD

Apparatus and stimuli.—A standard Color i and 21 colors[2] ($j = 1, 2, \ldots, 21$) were presented simultaneously on a rectangular board, 45 × 95 cm. Each of the 21 colors, .7 × 1.0 cm., was pasted on a piece of thick cardboard, 1.3 × 3.1 cm. The remaining part of the card was covered with gray paper (N 5). When the card was handled, the gray part was always picked up with a pincette which was covered by gray cloth. The standard color was also pasted on a thick cardboard, 1.0 cm. wide and 45 cm. long. This was always placed at the left side of the board. The *S* was seated in front of a table on which was the board. Gray cloth was

---

[2] The colored papers were obtained from Murakami Color Research Laboratory.

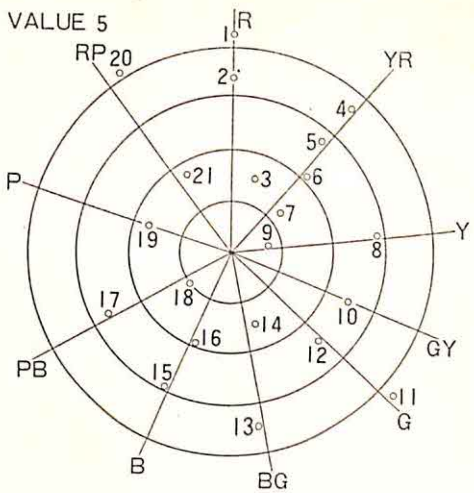


FIG. 1. Color stimuli in terms of Munsell hue and chroma. Concentric circles denote Chroma 2, 4, 6, and 8 respectively. Radius vectors are only for the sake of facilitating identification of colors.

hung from the ceiling around the table. The same gray cloth was pasted on the board to cover its entire surface. Three fluorescent 20-w. daylight tubes (Toshiba FL-SDL) evenly illuminated the surface of the board to 500 lux. Since the spectral energy distribution of this light source was known and the spectral reflectances of the colored papers were measured, it was possible to compute tristimulus values for the colors under this illumination. The Munsell designations were obtained from the tristimulus values through "the table of ICI equivalents of the recommended Munsell notation" (Newhall et al., 1943). Colorimetric data of the gray cloth were as follows; $x = .305$, $y = .302$, $Y = 18.0$, and 3.0P 4.8/1.0. A plot of the stimuli according to hue and chroma appears in Fig. 1 where radius vectors are introduced, just for the sake of convenience, to compare this configuration with those of the final results. In the Munsell system, for instance, 5Y is defined to represent the yellow as a major hue. Nevertheless, the Vector Y in Fig. 1 is not located at 5Y but at approximately 7Y. The Munsell values of the stimuli were between 4.6 and 5.0.

Procedure.—On the gray board, a standard Color i (45 cm. long, 1.0 cm. wide) was placed along the edge of the left side (45 cm.). Parallel to the 95-cm.-long side, 21 rows were marked by gray threads on the board, and a Color j was placed in each row at its left end. That means, 21 colors made a column next to the standard Color i on its right side. The same one as the standard was included among the 21 colors. The S was told that this duplicate of the standard would always be present and that he should place it right beside the standard. Then S was required to move, one by one, each of the other 20 colors to the right in the row according to the difference he observed between Color i and Color j. For the purpose of furnishing S with a unit for moving the colors, a pair of gray cards, either N 5.0 and N 5.5 or N 5.0 and N 4.5 were placed on the top of the board. The interval between the cards (.7 × 1.0 cm. each) was fixed at 3 cm. Should the difference between this pair of grays be represented by this distance (center to center), how far should the Color j be moved to the right to represent the difference between the Colors i and j? This was what S was instructed to do (with full knowledge that distances would be measured from center to center in the strips), hence the method may be called the multiple-ratio judgment. Usually, after having moved all the colors, S made subsequent changes in the positions of the colors by surveying the configuration of the colors made on the board as a whole. The final position of each color was thus found by a series of successive approximations. The same color as the standard was always placed by S at the left end of the row and from this point the placement were measured in centimeters. Occasionally S required to move a color beyond the right end of the board. A supplementary board was added in such cases. Each of the 21 colors took the place of the standard i in turn. The spatial order of the 21 colors in the column was randomized from trial to trial.

Subjects.—Four Ss participated in the experiment. All were normal in color vision. Id is the senior E and had served also as S in the previous experiment (Indow & Shiose, 1956). On the other hand, K, O, and M had neither experience in an experiment of this kind nor special knowledge about the Munsell system. K and O were women students in the introductory psychology course and M was a housewife. They found, however, no difficulty in converting the color differences into the spatial distances as required, after a few trials. And they were so naive that no special emphasis was necessary to have them take an unanalytical attitude in observing the color differences. As each of the 21 colors took the place of the standard i, one round consisted of 21 trials. Id repeated four rounds of the trials and the other Ss two rounds, respectively. In a sitting 10 to 14 trials were run.

Scaling procedure of the differences.—The first step of multidimensional scaling is to obtain scale values for the distances between all pairs of stimuli. It is important, however, that the subsequent multidimensional mapping is not independent of the nature of the scale on which the distances are determined. The distances have to be determined, in the sense of Stevens (1951), on a ratio scale and not on an interval scale, because the choice of an arbitrary origin for the scale will affect the dimensionality as well as the nature of the space to account for these interstimulus distances (Messick & Abelson, 1956; Torgerson, 1952, 1958). For obtaining the distances, Torgerson (1952, 1958) employed the complete method of triads, i.e., an extension of the method of paired comparison, and Messick (1954) used the method of successive intervals of a modified form. Either of the methods yields only an interval scale. Hence, they had to convert the relative distances thus obtained to the absolute ones by estimating the value of the additive constant. To choose the correct additive constant is not an easy task and its analytical solution is extremely tedious in computation (Messick & Abelson, 1956). In the previous experiment Indow and Shiose (1956) had to arrive at the absolute distances with the help of the estimated additive constant. Now this difficulty can be overcome by using the method of the present experiment, since the color differences are obtainable from the beginning on a ratio scale.

As the result of the placement of a Color j by S, a spatial Distance x_{ij} is obtained which should represent the observed difference between the Color j and the standard Color i in terms of an arbitrary Unit e_i. The distances x_{ij} $(j = 1, 2, \ldots, n)$ may be thought of as being expressed by the same Unit e_i because they are determined at a trial as the results of simultaneous comparisons with each other. However, in general, the Unit e_i will not remain constant if the standard Color i varies. Although a pair of grays are given as a guide for converting the color differences into the spatial distances, as a matter of fact, the proportional constant of the conversion will be more or less affected by the total relations between the standard Color i and the Color j $(j = 1, 2, \ldots, n)$. The differences, in general, may be relatively small with a standard color and relatively large with another. In other words, the proportionality will be kept constant among x_{ij}'s in so far as i is fixed, but not necessarily if i varies. Let us denote by δ_{ij} the color difference between i and j. Hence,

$$x_{ij(i)} = \delta_{ij} e_i \qquad [1]$$

The suffix in the parenthesis should be understood hereafter as indicating the unit in terms of which δ is quantified as x. Let us arrange $x_{ij(i)}$ in the form of the Matrix X. Then, taking arbitrarily kth row as a reference, let us consider the transformation

$$x_{ij(k)} = x_{ij(i)} c_{ik} \qquad [2]$$

where $x_{ij(k)}$ means δ_{ij} defined in terms of the unit of the kth row of X and by definition,

$$x_{ij(k)} = \delta_{ij} e_k = \delta_{ij} e_i \frac{e_k}{e_i} = x_{ij(i)} \frac{e_k}{e_i}$$

Hence, remembering $\delta_{ik} = \delta_{ki}$, c_{ik} will be given as follows,

$$c_{ik} = \frac{e_k}{e_i} = \frac{\delta_{ki} e_k}{\delta_{ik} e_i} = \frac{x_{ki(k)}}{x_{ik(i)}} \qquad [3]$$

where $x_{ki(k)}$ denotes the ith element in the kth row and $x_{ik(i)}$ the kth element in the ith row of X. The transformation [2] will be possible in as many ways as n, the number of rows of X. Taking the average

$$\bar{x}_{ij} = \frac{1}{n} \sum_k x_{ij(k)} = x_{ij(i)} \left(\frac{1}{n} \sum_k c_{ik} \right) = x_{ij(i)} \bar{c}_i \qquad [4]$$

$(c_{ii} = 1)$

and on account of Equation 3,

$$\bar{c}_i = \frac{1}{n} \sum_k \left(\frac{e_k}{e_i} \right) = \frac{\bar{e}}{e_i} \qquad [5]$$

where $\bar{e}$ is the average of e_i's. From Equations 1, 4, and 5, we have

$$\bar{x}_{ij} = \delta_{ij} \bar{e} \qquad [6]$$

and the whole observed color differences can be defined by a single common Unit $\bar{e}$. Due to random errors, $\bar{x}_{ij}$ will be somewhat different from $\bar{x}_{ji}$. Hence, the mean of $\bar{x}_{ij}$ and $\bar{x}_{ji}$, i.e., d_{ij}, was used through the present study as the value of δ_{ij} on the ratio scale.

As stated previously, Id made four repetitions on each of $x_{ij(i)}$, and the remaining Ss made two repetitions. The average was taken as the element of the Matrix X, and d_{ij} was obtained by the procedures described above for each S respectively.

To see how small the discrepancy is in general between $\bar{x}_{ij}$ and $\bar{x}_{ji}$ will serve as a criterion of consistency in the placements made by S. The correlation coefficients between the two were .98 for Id, .89 for K,

.93 for O, and .85 for M. Besides, the plots of $\bar{x}_{ij}$ with $\bar{x}_{ji}$ showed in all cases that the points plotted were scattered along the line passing through the origin with the slope of one. There was a high degree of consistency in each *S* to express δ_{ij} as either $x_{ij(i)}$ or $x_{ji(j)}$. The criterion tells us nothing, however, concerning whether the relationship between x and δ is really as simple as given in Equation 1. Even if δ_{ij} is replaced by $f(\delta_{ij})$, a function of δ_{ij}, the whole procedure of the scaling still holds and, instead of Equation 6, $f(\delta_{ij})\bar{e}$ will be obtained as $\bar{x}_{ij}$. In this case also the criterion will be met since $f(\delta_{ij})$ is equal to $f(\delta_{ji})$. It is a matter of assumption that the observable x is put in Equation 1 as proportional to the latent variable δ. No empirical verification was possible in making this assumption.

The multidimensional scaling was applied to the data of Id and also to the pooled data of K, O, and M. The d_{ij}'s of the three *S*s were pooled as follows. In order to reconcile the difference in the Unit $\bar{e}$, the d_{ij}'s of each *S* were plotted against the d_{ij}'s of Id and the former were multiplied by a constant so that the points plotted were scattered along the line passing through the origin with the slope of one. By taking the unit of the scale of Id as reference, the scales of all *S*s were thus defined by a common standard. The converted d_{ij}'s of the three *S*s were averaged to give the Common Distance Matrix *D*.

Multidimensional scaling.—Let i and k be alternative subscripts for stimuli (i, $j = 1, 2, \ldots, n$) and let b_{ij} be the scalar product of the vectors from an origin to Points i and j in a real Euclidean space of r dimensions. As suggested by Torgerson (1952, 1958), it is convenient to place the origin at the centroid of all of the stimuli. Let the scalar product thus defined be denoted by b^*_{ij}, then b^*_{ij} can be derived from Matrix *D* by the formulae given by Torgerson. Once Matrix B^*, whose elements are b^*_{ij}, is obtained, it becomes possible to determine the dimensionality of the space and to obtain a_{ip}, the projections of points on an arbitrary set of orthogonal coordinate axes, because B^* can be written as follows:

$$B^* = AA' \qquad [7]$$

where $A = (a_{ip})$ and $p = 1, 2, \ldots, r$. A' is the transposition of A. It is important to note that Equation 7 does not hold unless B^* is positive semidefinite, that means, r latent roots of B^* have relatively large positive values and the remaining roots vanish except for errors. In other words, the colors will be considered as lying in a real Euclidean space of r dimensions if B^* is shown to be of rank r and to have no large negative latent root.

Two B^*s were constructed for Id and for K, O, and M. In accordance with Equation 7, B^* was factored by the centroid method in the case of Id and by the multiple-group method (Harman, 1954) in the other case. Three factors were extracted in both cases. For the following reasons, however, the third factor could be discarded and we could conclude that B^* was of rank two, except for variable errors, for Id as well as for K, O, and M: (*a*) The second factor residuals all seemed sufficiently small; (*b*) The contribution of the additional factor was relatively small. The ratio $\Sigma_i a^2_{i3}/\Sigma_i a^2_{i1}$ was .13 for Id and .17 for K, O, and M. (*c*) The criterion for B^* to be of rank two is given by Torgerson as follows:

$$\sum_i \sum_j (b^*_{ij})^2 = (\sum_i a^2_{i1})^2 + (\sum_i a^2_{i2})^2 + 2(\sum_i a_{i1}a_{i2})^2 \qquad [8]$$

In the case of Id, 81.18×10^7 was obtained for the left-hand term and 76.70×10^7 for the right. The coincidence seems close enough to discard the third factor, though it is a little worse than the coincidence in Torgerson's experiment (1958). The ratio between the left- and the right-hand members of the equation was 1 to .944 in the former and 1 to .981 in the latter. For the data of K, O, and M, the criterion would be inappropriate because of the method of factoring. In order to apply the criterion, it is desirable that the a_{ip}'s in the equation be the results of the method of principal axes or, at least, of the centroid method. (*d*) No particular relationship was found between a_{i3} and the fluctuation of the stimuli in Munsell value.

RESULTS AND DISCUSSION

The configurations given by *A*s are shown in Fig. 2 and Fig. 3 where the coordinate axes are omitted. Instead, radius vectors are drawn so as to run through colors of the same hue and a series of curves is also formed for chroma. Since the scale units have been equated as stated before, the two configurations can be directly compared with each other.

It is interesting to observe that

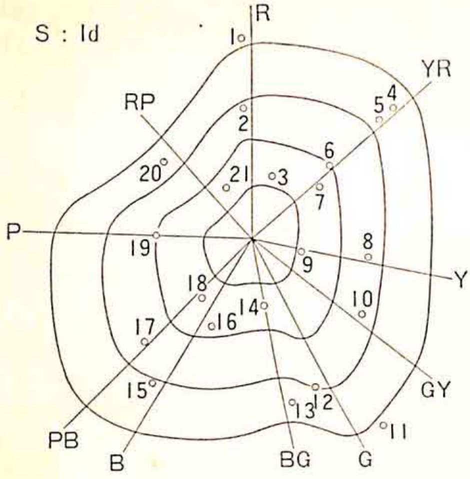


FIG. 2. Two-dimensional mapping of colors for Subject Id. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration obtained.

the colors of the same Munsell hue appear along the line segments in the order corresponding to their chroma and, besides, all line segments converge into a single point. Furthermore, the pairs of the complementary hues are identified by the radius vectors lying more or less in the opposite direction to each other. Except for B and PB in Fig. 3, the order of hue is exactly what is expected. As to the spacing, however, the disagreement with the Munsell notation is conspicuous. Nevertheless, the patterns in Fig. 2 and Fig. 3 coincide closely with each other in the interval from one hue to the next. The exceptions are the position of G between GY and BG, and the positions of PB and B between BG and P. Comparison with Fig. 1 shows that the two spacings of hue obtained deviate more or less in the same way from the Munsell notation. For example, the interval between R and YR and also that between YR and Y are too large and the interval between B and PB is too small in the results of this study.

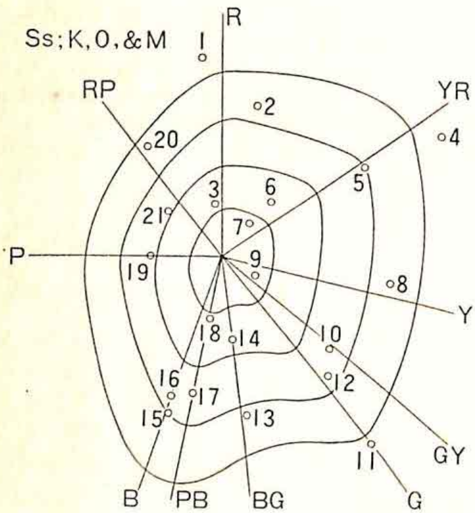


FIG. 3. Two-dimensional mapping of colors for *S*s K, O, and M. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration obtained.

The equichroma curves in Fig. 2 and 3 were drawn so that, though the interval between the neighboring curves may differ from hue to hue, the interval on the same hue should be the same irrespective of chroma. If the series of curves in Fig. 2 and 3 are superimposed directly on each other, they are in good agreement in the region from RP through Y to BG. On the contrary, in the region of B and P the disagreement is marked. The disagreement, however, is apparently due to the discrepancy in spacing of B and PB. There is no difference between Fig. 2 and 3 of the interval from one equichroma curve to the next along each of these hues. It might be well to consider whether a set of concentric equichroma circles, as defined in the Munsell notation, can also be fitted to the data. Excepting the region around RP, the fitting was not entirely impossible and a plot of the chroma of the colors thus defined with the Munsell chroma of

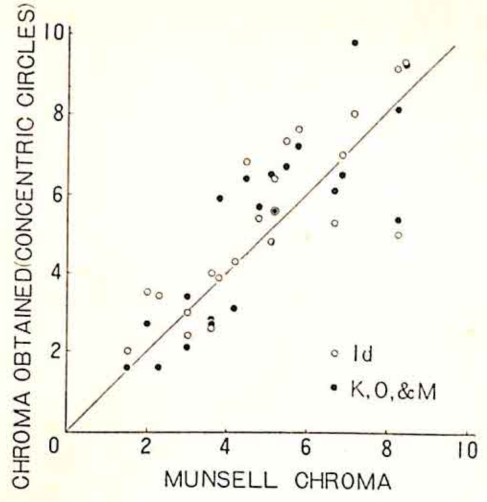


FIG. 4. Munsell chroma with chroma defined in terms of concentric equichroma circles fitted to the configurations in Fig. 2 and 3.

the stimuli is given in Fig. 4. The root mean square of the discrepancies between the two kinds of chroma was 1.24. If the restriction that equal intervals in Munsell chroma should correspond to equal linear extents irrespective of hue is withdrawn, obviously the goodness of fit improves. When the chroma in terms of the equichroma curves in Fig. 2 and 3 was taken as the ordinate, the scatter of the points was so reduced that the root mean square of the discrepancies decreased to 7.6.

Setting aside the question as to the adequacy of the spacing of the Munsell colors, it is satisfactory to observe that the colors varying in hue and chroma were fitted by the configuration in two-dimensional Euclidean space. If uniformity of the scale *in the large* is taken into consideration, among the basic postulates of the Munsell system the most questionable one will be that, of the polar coordinate in a Euclidean plane of constant lightness, hue is represented by the magnitude of the argument and chroma by the length of the radius vector. In order that the model may hold good, the difference perceived between two colors not of the same hue, say a and b, has to vary in a certain manner, if the chroma of each color is changed. Denoting by θ the difference between the two arguments corresponding to Hues a and b and by ρ_1 and ρ_2 the lengths of the two radius vectors representing the chromas of the pair of colors, the difference will be proportional to the distance

$$\sqrt{\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2 \cos\theta} = \Delta$$

Let us consider, then, a pair of colors which are of Hues a and b respectively, but the chroma of each color is n times in magnitude that of the corresponding member of the former pair. By definition, the difference between the latter pair has to be proportional to the distance

$$\sqrt{n^2\rho_1^2 + n^2\rho_2^2 - 2n^2\rho_1\rho_2 \cos\theta} = n\Delta$$

That means, the difference between a pair of colors has to increase n times if both of the chromas of the members are increased n times. Otherwise, the model will not hold good.

It must be remembered that the configuration under discussion is more than a mere schema, because the points are located in such a way that the distances between all pairs of the points are proportional to the observed differences between the corresponding colors, no matter how large or small the differences are. Let s_{ij} denote the distance between the Points i and j in the multidimensional configuration. When the agreement between s_{ij}, i.e., the difference in theory, and d_{ij}, i.e., the observed difference scaled, was shown in terms of the correlation coefficient, r was .96 for Id. The multidimensional mapping for K, O, and M in Fig. 3 was made from the pooled data. Nevertheless, using s_{ij} determined from Fig. 3, r was .84 for K, .88 for O, and .85 for M. No qualitative individual differences could be observed among *S*s. However, the correlation coefficient will not be sufficient to tell the exact relationship existing between s_{ij} and d_{ij}. In the case of K, O, and M, as expected, the points plotted were scattered along the line passing the

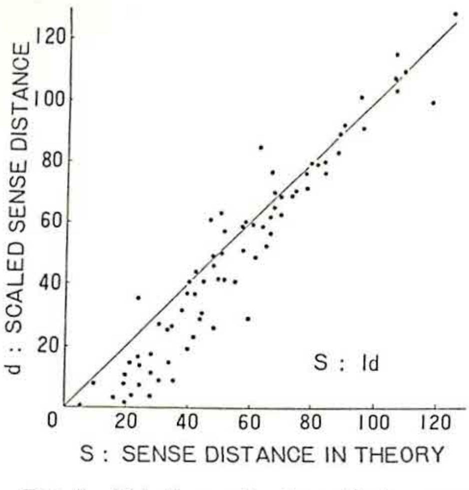


FIG. 5. Relation of theoretical sense-distances to scaled differences between all possible pairs of the colors. A third of all the pairs drawn randomly are plotted. Result for Subject Id. Theoretical sense-distance is taken from Fig. 2.

origin with the slope of one. As shown in Fig. 5 the regression of d_{ij} on s_{ij} is slightly curved, in the case of Id. The tendency that $d_{ij} < s_{ij}$ is predominant where $s_{ij} < 50$. It was mentioned previously that the distance $x_{ij(i)}$ may not be proportional to the perceived difference δ_{ij} between Colors i and j. Instead, it may be proportional to a function of δ_{ij}. The phenomenon observed in Fig. 5 may be ascribed partially to the non-proportionality, if any, existing between the observable x and the latent variable δ.

SUMMARY

The difference perceived between the standard color and each of 21 colors was converted into spatial distance by moving the latter one by one. As each of the 21 colors took the place of the standard, a matrix was obtained whose elements were the ratio scales of the differences between all possible pairs of the 21 colors. The adjustment was made with respect to the difference, if any, in the unit of converting, from trial to trial and from *S* to *S*. Then, the method of multidimensional scaling was applied to the matrix. The colors varied in hue and saturation over the widest possible range but were kept constant in lightness at 5 in Munsell value. Four *S*s participated and the data of the individual *S* (Id) and the pooled data of the remaining *S*s (K, O, and M) were analyzed separately. In both cases, the two-dimensional configuration obtained in a Euclidean space reproduced with high fidelity the structure inherent in the color differences perceived by each *S*. The colors of the same Munsell hue appeared along a line segment in the order corresponding to their chroma, and all line segments converged into a single point at the center. With one exception (PB in K, O, and M), the order of hue was exactly as expected. Hence, it may be concluded that one of the basic postulates of the Munsell system was experimentally confirmed even when color differences *in the large* were taken into consideration. As to the spacing of colors, however, the disagreement with the Munsell notation was observed, especially with respect to hue.

REFERENCES

GODLOVE, I. H. Improved color-difference formula, with applications to the perceptibility and acceptability of fadings. *J. Opt. Soc. Amer.*, 1951, **41**, 760–772.

HARMAN, H. H. The square root method and multiple group methods of factor analysis. *Psychometrika*, 1954, **19**, 39–55.

INDOW, T., & SHIOSE, T. An application of the method of multidimensional scaling to perception of similarity or difference in colors. *Jap. psychol. Res.*, 1956, **3**, 45–64.

JUDD, D. B. *Color in business, science, and industry.* New York: Wiley, 1952.

MESSICK, S. J. The perception of attitude relationships: A multidimensional scaling approach to the structuring of social attitudes. Unpublished doctoral dissertation, Princeton Univer., 1954.

MESSICK, S. J. An empirical evaluation of multidimensional successive intervals. *Psychometrika*, 1956, **21**, 367–375.

MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21**, 1–15.

NEWHALL, S. M. Preliminary report of the OSA subcommittee on the spacing of the Munsell colors. *J. Opt. Soc. Amer.*, 1940, **30**, 617–645.

NEWHALL, S. M., NICKERSON, D., & JUDD, D. B. Final report of the OSA subcommittee on the spacing of the Munsell colors. *J. Opt. Soc. Amer.*, 1943, **33**, 385–418.

RICHARDSON, M. W. Multidimensional psychophysics. *Psychol. Bull.*, 1938, **35**, 659. (Abstract)

SHIOSE, T. An application of the method of multidimensional scaling to the perception of similarity or difference in colors. II. *Jap. J. Psychol.*, 1958, **28**, 375–385. (English summary)

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

TORGERSON, W. S. Multidimensional scaling: Theory and method. *Psychometrika*, 1952, **17**, 401–419.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

(Received June 8, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

MULTIDIMENSIONAL MAPPING OF MUNSELL COLORS VARYING IN HUE, CHROMA, AND VALUE[1]

TAROW INDOW AND KEI KANAZAWA

Keio University

Evidence in support of the basic postulates of the psychological color solid has been accumulated by the application of the multidimensional scaling. Torgerson (1958), Messick (1954), and Indow and Shiose (1956) showed that Munsell colors of the same hue (5R) could be fitted by a configuration in a two-dimensional Euclidean space, and that the orthogonal coordinates were identified as representing Munsell value (lightness) and chroma (saturation) respectively. Besides, the two-dimensional mapping of colors of constant lightness (Value 5) gave the following results (Indow & Uchizono, 1960). The colors lay in a Euclidean plane and each hue was represented by one of the radius vectors which converged into a single point corresponding to gray. Around this point, it was possible to define the series of equichroma curves. The hypothesis underlying the construction of an ideal color solid, e.g., the Munsell system, would be that equal sense distances shall correspond to equal linear extents in a Euclidean space, no matter how large or small the magnitude of the sense distances may be (Balinkin, 1939; Newhall, 1939). Since the multidimensional mapping is carried out on the basis of sense distances between every possible pair of colors employed, the findings mentioned above seem to provide experimental evidence in support of at least the fundamental hypothesis. As a matter of course, there are minor discrepancies between the results of the experiments and the Munsell spacing of color. The discrepancies will, however, be removed by adjusting the locations of the Munsell samples, if necessary. On the contrary, if the hypothesis were shown untenable, the color solid could not be more than a mere schematic representation.

In the previous experiments, the variation among colors has been more or less restricted. Colors have been used which differ from each other in only two attributes. Shiose (1958) made an experiment with the colors differing in value, chroma, and hue. However, the hue ranged only from 6R to 6YR. Messick (1956) reconsidered the results of his experiment taking into consideration slight hue variation of the colors used (roughly from 2R to 5.5R), and found that the variation could be accounted for by the addition of the third dimension. In the present experiment, the tridimensional mapping of colors over a considerable range was attempted. So far, no qualitative individual difference has been found in the multidimensional configuration constructed separately for each *S* (Indow & Shiose, 1956; Indow & Uchizono, 1960). Hence, the multidimensional scaling was applied to the pooled data of 4 *S*s participating in the

---

[1] This research was supported by a grant from Ministry of Education, Japan. The investigators are indebted to S. Ura. Under his supervision a part of the computations involved in this research was carried out by the electronic automatic computor in Japan Electronic Industry Development Association. Appreciation is also expressed to M. Matthewson for her helpful advice concerning English. Thanks are due to T. Uchizono for her help in the preparation of this report.

present study to give a common configuration.

METHOD

Apparatus and stimuli.—The viewing conditions and the manner of presenting the stimuli on a gray board were exactly the same as those in the previous experiment (Indow & Uchizono, 1960). The instructions given to *S* were also the same; namely, the perceived difference between a standard Color *i* and each of the Colors *j* ($j = 1, 2, \ldots, 24$) was to be converted into spatial distance by the method of multiple ratio judgment. Tristimulus values for the 24 colors under the given illumination were calculated from the spectral reflectances of the colored papers and the spectral energy distribution of the light source. Then, their equivalents in the recommended Munsell notation are obtained. The colors were of two levels in lightness. The brighter ones were from 4.3 to 4.9 and the darker ones were from 6.4 to 7.3 in terms of Munsell value. For convenience, the former will be called the stimuli of Value 5 and the latter the ones of Value 7 in the figures. A plot of the stimuli according to the hue and chroma appears in Fig. 1 where arbitrary radius vectors are given only for the sake of convenience. In other words, the hue names in the figure do not necessarily mean the corresponding principal hues in the Munsell notation.

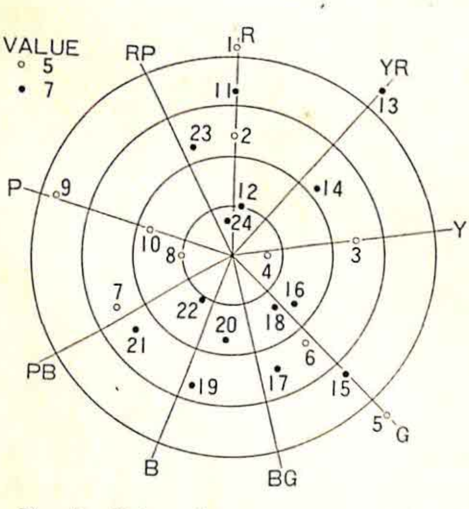


FIG. 1. Color stimuli in terms of Munsell hue and chroma. Concentric circles denote Chroma 2, 4, 6, and 8 respectively. Radius vectors are only for the sake of facilitating identification of colors.

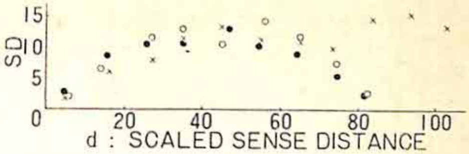


FIG. 2. *SD*s of scaled distances for four *S*s (×) and *SD*s of three repeated judgments for Subject Id (○) and for T (●) as functions of magnitude of the sense distance scaled.

Subjects.—All *S*s were normal in color vision. Id and Sh were investigators of the previous study (Indow & Shiose, 1956), while T and Si were students of Keio University who had neither experience in an experiment of this kind nor special knowledge of the Munsell system. Each *S* made three repetitions on each pair of Colors *i* and *j*.

Scaling procedure of sense distances.—As to the three spatial distances made by each *S* on the board to represent the perceived difference between the pair of Colors *i* and *j*, the average was obtained. To the set of the averages the scaling procedure was applied that was described in detail in the previous article (Indow & Uchizono, 1960). Thus, the sense distances were scaled once separately for each *S* to give $d'_{ij(\alpha)}$ where the subscript α denotes *S*. The next step was to equate e_α, the unit of the scale which may differ from *S* to *S*. Taking the unit of the scale of Id as standard, each of the remaining e_α's was multiplied by a constant so that a plot of the adjusted $d_{ij(\alpha)}$ against $d_{ij(\text{Id})}$ was scattered along the line passing the origin with the slope of one. The scaled values expressed in terms of the common Unit e_{Id} were then averaged. The averages, d_{ij}'s, constituted a Common Distance Matrix *D* for the four *S*s. The multidimensional scaling procedures were applied to *D*.

The *SD*s of the four values combined, i.e., $d_{ij(\alpha)}$'s and $d_{ij(\text{Id})}$, are plotted as a function of the magnitude of d_{ij} in Fig. 2. For the sake of comparison, the *SD*s of the three repeated judgments made by Id and also by T are shown in the same figure. The heterogeneity among *S*s was of the same magnitude as the fluctuations within the repeated responses of an individual *S*. It is interesting to observe that the errors are minimum at the lowest and at the highest values in the case of the individual data. This is the pattern that is often encountered in psychological tests (Mollenkopf, 1949).

Multidimensional scaling.—From Distance Matrix *D*, Matrix B^* was obtained which consists of b^*_{ij}, the scalar product of the vec-

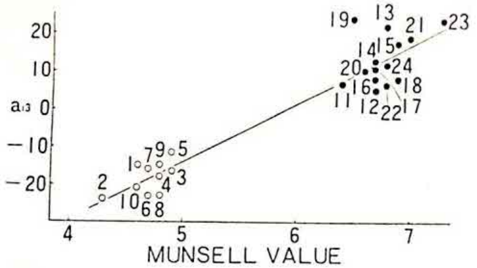


FIG. 3. Munsell value with projection of the points on third axis which is supposed to represent lightness of color.

tors to Points i and j from the centroid of the 24 stimuli used. Matrix B^* was then factored to give Matrix A, by means of obtaining the latent vectors and the latent roots of B^*. All the computations were carried out by the electronic automatic computor. Matrix A gave a mapping of the 24 colors in a Euclidean space of three dimensions. Matrix B^* was warranted of rank 3, since the third factor residuals were sufficiently small and the criterion given by Torgerson:

$$\sum_i \sum_j (b^*_{ij})^2 = \sum_p (\sum_i a^2_{ip})^2 + 2 \sum_{p \neq q} (\sum_i a_{ip}a_{iq})^2$$

was met when p was 3. The right- and left-hand members of the equation were 11.03×10^8 and 10.77×10^8, respectively. The ratio between the two is 1 to .976, whereas that in the study of Torgerson (1958) was 1 to .981. It was discovered that the fit to the Munsell notation became better when the coordinate axes were rotated. The projections of the points on each of the three orthogonal axes after the rotation are used as a_{ip}'s. As will be shown later, the first and second axes determine the plane of constant lightness, while the third corresponds to Munsell value.

RESULTS AND DISCUSSION

A plot of a_{i3}'s with the Munsell value of the stimuli is shown in Fig. 3. The separation of the two levels of lightness is perfect and, besides, the fluctuations in the Munsell value within each level is also somewhat reflected in the results, because the scatter within the level gives a correlation coefficient of .51 in the darker level and .40 in the brighter level.

The configuration in terms of hue and chroma is given in Fig. 4. The colors of the two lightness levels are plotted together. Radius vectors running through colors of the same hue are drawn and a series of curves is also formed for chroma on the same principle as in the previous study (Indow & Uchizono, 1960). The results are substantially in agreement with the results of the previous experiment. It is interesting that the configurations, in the brighter and darker levels, could be projected into a single plane without contradiction. For instance, YR in the brighter level comes between R and Y in the darker level, and Colors 15, 16, and 18 of the brighter level form a radius vector with 5 and 6 of the darker level in accordance with the definition of stimuli in Fig. 1.

In the previous study (Indow & Uchizono, 1960), two 2-dimensional configurations of a constant lightness were constituted, for individual data of Id and for the pooled data of K, O, and M, and no remarkable qualitative difference was found between the two. Hence, if the two configurations are combined into one, comparison between the results of the previous and present studies will be

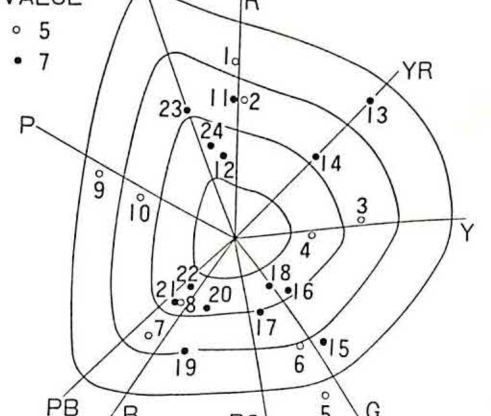


FIG. 4. Two-dimensional mapping of colors in terms of hue and chroma. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration obtained.

facilitated. Combining two configurations is possible by the method of least squares.

Let us suppose that two configurations are given, both of which are represented in polar coordinates, as (ρ_{1i}, θ_{1i}) and (ρ_{2i}, θ_{2i}), where ρ denotes a radius vector and θ an argument with reference to an arbitrary axis. Then, the problems are to find the angle which the two reference axes should form, in order to give the closest agreement between the configurations, and also to find the coefficient π which should adjust the difference, if any, between the units of two scales. Denoting by α the angle between the reference axes of two coordinates, and by a_i the distance between a Point i in a configuration and its corresponding point in another, then

$$a_i^2 = \rho_{1i}^2 + \pi^2\rho_{2i}^2 - 2\pi\rho_{1i}\rho_{2i}\cos(\theta'_{2i} - \theta_{1i}) \qquad [1]$$

where θ'_{2i} represents the argument of a Point i in the second configuration expressed in terms of the first polar coordinate, i.e.,

$$\theta'_{2i} = \theta_{2i} + \alpha$$

and the unit of the second is equated to that of the first.
Defining

$$A = \sum_i \rho_{1i}\rho_{2i}\cos(\theta_{2i} - \theta_{1i})$$

$$B = \sum_i \rho_{1i}\rho_{2i}\sin(\theta_{2i} - \theta_{1i})$$

the quantity which is to be minimized is given by

$$Q = \sum_i a_i^2 = \sum_i \rho_{1i}^2 + \pi^2 \sum_i \rho_{2i}^2 - 2\pi(A\cos\alpha - B\sin\alpha) \qquad [2]$$

Differentiating Equation 2 with respect to α, we have

$$\frac{\partial Q}{\partial \alpha} = 2\pi(A\sin\alpha + B\cos\alpha) \qquad [3]$$

The minimizing value $\hat{\alpha}$ is found by equating Equation 3 to zero. Simplifying the result gives

$$\hat{\alpha} = \tan^{-1}\frac{-B}{A} \qquad [4]$$

Differentiating Equation 2 with respect to π, we have

$$\frac{\partial Q}{\partial \pi} = 2\pi\sum_i \rho_{2i}^2 - 2(A\cos\alpha - B\sin\alpha) \qquad [5]$$

The minimizing value $\hat{\pi}$ is found by equating Equation 5 to zero. By using Equation 4, we have

$$\hat{\pi} = \frac{1}{\sum_i \rho_{2i}^2}(A\cos\hat{\alpha} - B\sin\hat{\alpha}) \qquad [6]$$

The two-dimensional configuration of K, O, and M (Indow & Uchizono, 1960) was superimposed upon that of Id (Indow & Uchizono, 1960) in this way ($\hat{\pi} = .94$, $\hat{\alpha} = 3° \, 30'$, taking R as the reference axis). Taking the middle point of a_i, the combined configuration of the colors was developed, which is shown in Fig. 5 where the numbering of the colors is to be understood as defined in Fig. 1 in Indow and Uchizono (1960). Radius vectors and a series of equichroma curves are formed on the same principles as before. Comparison with Fig. 4 shows that agreement between the results of the two experi-

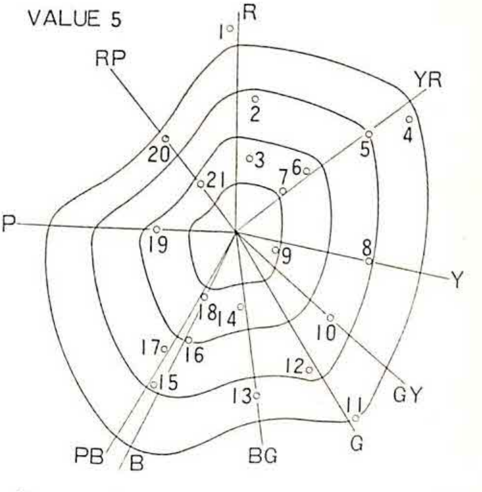


FIG. 5. Two configurations obtained by Indow and Uchizono (1960) are combined into one in order to be compared with Fig. 4. Number of colors corresponds to that given in Fig. 1 of Indow and Uchizono. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration shown here.

ments is satisfactory, except for the definition of chroma in the region of RP and the spacings of P and RP. As a matter of course, Fig. 5 can be superimposed upon Fig. 4 by the method described above. In so doing, it may be appropriate to take as references the points of intersection of the equichroma curves with the radius vectors used in common in both investigations. Taking R as the reference axis, it was found that the closest fit was obtained by $\hat{\pi} = 1.02$ and $\hat{\alpha} = -2° 52'$. Should loci of constant hue and constant chroma be defined as averages of the results of the two experiments, the equichroma curves do not deviate so much from the concentric circles, and the loci of constant hue are substantially in agreement with the Munsell spacings except for PB. Radius Vector PB appeared too close to B in both of the two experiments. It is the impression of the investigators that, looking at the Munsell Book of Color (Pocket Edition), people seem unanimous in pointing out that the size of the step from one sheet of constant hue to the next is relatively small in the region from B to PB.

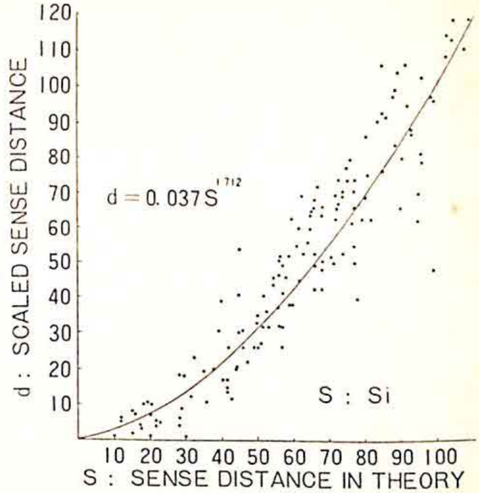


FIG. 6. Relation of theoretical sense distances to scaled differences between all possible pairs of the colors. A third of all the pairs drawn randomly are plotted. Result for Subject Si. Theoretical sense distance is taken from Fig. 3 and 4.

Let s_{ij} denote the distance between the points i and j in the three-dimensional configuration of the present study. When the agreement between s_{ij}, the color difference in theory, and the adjusted $d_{ij(\alpha)}$, the scaled sense distance for the subject α, is shown in terms of the correlation coefficient, r was .94 for Id, .90 for Sh, .94 for T, and .95 for Si. Obviously the multidimensional configuration of the colors reproduces with fidelity the structure inherent in the color differences perceived by each *S*. Nevertheless, it was discovered that the regression of $d_{ij(\alpha)}$ upon s_{ij} is slightly curved. While the curvilinear relationship was observed only in Id among *S*s participating in the previous experiment (Indow & Uchizono, 1960), it was conspicuous in all of the four *S*s in the present experiment. The relationship observed in Id in this experiment was exactly the same as that in the previous one. In Fig. 6 is given a typical example of the relationship observed in the remaining *S*s that seems to be more or less parabolic in all the cases. The empirical equation in Fig. 6 was fitted by inspection. The phenomenon seems to be a subject eminently worthy of theoretical study.

When the chroma of the colors defined in terms of the equichroma curves in Fig. 4 was plotted against the chroma in the Munsell notation of the stimuli, it became apparent that the fit was not satisfactory. The root mean square of the discrepancies was .97 and this was not so large compared with .76 in the previous study. Nevertheless, the tendency was apparent that the points systematically deviated above the line in the region of the lower chroma. Hence, the size of the step from Chroma 0 to 2 was defined differently from the size of the other steps along the same vector that was kept constant. The equichroma curves thus defined are given in Fig. 7. When the chroma in terms of these curves was plotted against the chroma in the Munsell notation, the above mentioned tendency disappeared and the goodness of fit improved as shown by the fact that the root mean square of the discrepancies was .73. It is to be

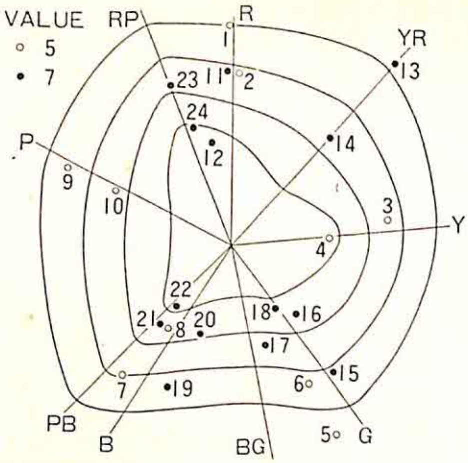


FIG. 7. Two-dimensional mapping of colors where equichroma curves are formed on the principle that size of the step from Chroma 0 to 2 may differ from sizes of the other steps which are kept constant in each hue.

noted that the equichroma curves approximate concentric circles on this principle. The loci of Chroma 2 were irregular in shape. Bellamy and Newhall (1942) determined limens with respect to chroma from achromatic origin in the five principal hues. The limens were given in terms of the Munsell scales. No particular relation was found, however, between the size of the first chroma step in the present experiment and that of the stimulus limen of chroma in the experiment of Bellamy and Newhall. It is most likely that there is a causal connection between this phenomenon, i.e., that the first chroma step was wider, and the curvilinear relationship between *s* and *d*. If the multidimensional scaling is applied to the absolute interstimulus distances which are obtained by adding a constant to relative interstimulus distances, the choice of a particular value of the additive constant will affect in theory the subsequent multidimensional mapping of the stimuli. Messick and Abelson (1956) showed that, if the additive constant is overestimated in a two-dimensional mapping, the configuration of stimuli becomes larger and "convex," and if underestimated, the configuration becomes smaller and "concave." It is true that, in the present study, the color differences have been obtained on a ratio scale from the beginning to make the estimation of the additive constant unnecessary, and that no curvature was discovered in the planes of constant lightness when a_{i3} was plotted against a_{i1}, or a_{i2}. We have no hypothesis at present to account for the problem under discussion. Certainly it awaits further inquiry.

In the Munsell system, the relative size of the units of the three attributes has been in dispute. For instance, Nickerson and Stultz (1944) consider that 1, 2, and 3 steps in V, C, and H respectively are visually equivalent in the region of 5C, and 1, 1, and 2 steps in the region of 3C, whereas, according to Bellamy and Newhall (1942), 1, 8, and 22 steps in V, C, and H respectively are defined as the same in perception in the region of 6C. On the other hand, 1 step in V is said to correspond to 4 steps in C by Godlove (1951). In the two-dimensional configurations individually constructed, Indow and Shiose (1951) obtained the following correspondences between the steps in V and in C; 1 to 3.1, 1 to 3.0, and 1 to 2.3 for the three *S*s respectively. The correspondence in the present study was 1 to 1.8, when the average of the sizes of the one-chroma steps in all the hues in Fig. 4 was compared with the size of the one-value step in Fig. 3. From Fig. 3 and 5, it may also be inferred that the correspondence was 1 to 2.0 in the average in the experiment by Indow and Uchizono (1960). In any case, the relative size of chroma step appears larger in these two experiments than in the previous experiment by Indow and Shiose. It is to be noted, however, that the colors used in the experiment by Indow and Shiose ranged in chroma from 4 to 12. If the correspondence is determined in the present experiment from Fig. 3 and 7, that is to say, if the sizes of the steps above Chroma 2 only are averaged, 1 step in V becomes equivalent to 2.7 steps in C, which is in agreement with the correspondence obtained in the experiment by Indow and Shiose.

SUMMARY

The sense distances between all possible pairs of the 24 colors were scaled by the method of multiple ratio judgment and by the procedures of adjusting the differences, if any, in the unit by which *S*s converted color differences into spatial distances. Then, the method of multidimensional scaling was applied. As the colors varied in the three attributes, the three-dimensional configuration of the colors was obtained for the pooled data of the four *S*s participating. The agreement among *S*s was satisfactory in spite of the fact that half of the *S*s were inexperienced. The axis corresponding to the Munsell value was orthogonal to the plane of constant lightness and the configuration of the colors in terms of Munsell hue and chroma was substantially in agreement with that obtained in the previous study. Both deviated somewhat in the same way from the Munsell notation, especially in the spacing of hue. It was pointed out that the equichroma curves fit the data better if the size of the chroma step from 0 to 2 is allowed to be wider than that of the remaining chroma steps. The correspondence between the steps of Munsell value and of Munsell chroma was 1 to 1.8 when the equichroma curves were drawn evenly in each hue. It becomes 1 to 2.7, however, with respect to the chroma steps, except for the first ones, if the latter are made wider than the former. It was discovered that the relation of the distances between all possible pairs of the points plotted in the three-dimensional configuration to the scaled sense distances from which the mapping was made was not linear but rather parabolic.

REFERENCES

BALINKIN, I. A. Industrial color tolerances. *Amer. J. Psychol.*, 1939, **52**, 428–448.

BELLAMY, B. R., & NEWHALL, S. M. Attributive limens in selected regions of the Munsell color solid. *J. Opt. Soc. Amer.*, 1942, **32**, 465–473.

GODLOVE, I. H. Improved color-difference formula, with applications to the perceptibility and acceptability of fadings. *J. Opt. Soc. Amer.*, 1951, **41**, 760–772.

INDOW, T., & SHIOSE, T. An application of the method of multidimensional scaling to perception of similarity or difference in colors. *Jap. psychol. Res.*, 1956, **3**, 45–64.

INDOW, T., & UCHIZONO, T. Multidimensional mapping of Munsell colors varying in hue and chroma. *J. exp. Psychol.*, 1960, **59**, 321–329.

MESSICK, S. J. The perception of attitude relationships: A multidimensional scaling approach to the structuring of social attitudes. Unpublished doctoral dissertation, Princeton Univer., 1954.

MESSICK, S. J. An empirical evaluation of multidimensional successive intervals. *Psychometrika*, 1956, **21**, 367–375.

MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21**, 1–15.

MOLLENKOPF, W. G. Variation of the standard error of measurement. *Psychometrika*, 1949, **14**, 189–229.

NEWHALL, S. M. The ratio method in the review of the Munsell colors. *Amer. J. Psychol.*, 1939, **52**, 394–405.

NICKERSON, D., & STULTZ, K. F. Color tolerance specification. *J. Opt. Soc. Amer.*, 1944, **34**, 550–570.

SHIOSE, T. An application of the method of multidimensional scaling to perception of similarity or difference in colors. II. *Jap. J. Psychol.*, 1958, **28**, 375–385. (English summary).

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

(Received June 8, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

RESISTANCE TO EXTINCTION AS A FUNCTION OF DEGREE OF REPRODUCTION OF TRAINING CONDITIONS[1]

MELVIN H. MARX

University of Missouri

The present study is one of a series designed to test some implications of the view that resistance to extinction is primarily a function of maintenance of the motivation of the animal to continue to perform the instrumental act, independently of any variation of the drive factor manipulated by incentive deprivation. In a previously reported experiment (Marx, 1958), animals extinguished with an empty foodcup present on half of the trials showed greater resistance to extinction than animals run with an empty foodcup present on every trial, following continuous reward during training.

The particular hypothesis tested in this experiment is that resistance to extinction is positively related to the degree of reproduction of training conditions. Hungry animals were trained in a runway with three distinctive endboxes, only one of which ever contained food. They were extinguished with only the formerly positive, only one of the formerly negative, both of the formerly negative, or all three of the endboxes present.

METHOD

Subjects

The *S*s were 36 experimentally naive female albino rats, approximately 120 days old at the start of the experiment.

Apparatus

A simple straight runway was used. It had a 1-ft. startbox, a 7-ft. stem, and a $1\frac{1}{2}$-ft. endbox (see Marx, 1958, for further details). Starting time, running time, and time in the endbox were all recorded automatically by means of floor switches. Black, white, and striped endboxes were used in training; a neutral Masonite endbox was used in accommodation.

Experimental Design

All *S*s were given the same accommodation and training. White, black, and striped endboxes were used in training, but each *S* found food only in one of these. Three *S*s in each group found food in each of the three endboxes. Extinction was run under four different endbox conditions (nine *S*s per group): Group P: formerly positive only; Group N: one of the formerly negative endboxes only; Group NN: both of the formerly negative endboxes; Group NNP: both of the formerly negative endboxes and the formerly positive endbox.

Procedure

Accommodation.—The *S*s were placed on a 23-hr. food deprivation schedule for five days prior to the start of accommodation to the apparatus. They were fed Little Friskie pellets for 1 hr. per day.

On the first day of accommodation, *S* was placed in the apparatus with all doors open, the neutral insert in the endbox, and no foodcup present. It was allowed to explore the runway and endbox for 15 min. A foodcup with five Little Friskie pellets was placed in the endbox and *S* was permitted

---

[1] This investigation was supported in part by a research grant, M-817, from the National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland. Thanks are due to Donald Morgan, who collected the data, and Aaron J. Brownstein, who performed the statistical computations and criticized the manuscript.

to eat these. Then *S* was removed from the apparatus and placed in an individual detention cage.

After 1 hr. *S* was returned to the apparatus and placed in the endbox by the foodcup, which contained two pellets. It was allowed to eat these and was then returned to the detention cage. After 40 sec. it was returned to the endbox just inside the closed door, at the end opposite to the foodcup. After *S* ate the two pellets in the cup, it was returned to the detention cage. Five additional trials were given at intervals of 40 sec. in the same manner as described above except that *S* was placed progressively further away from the endbox. The final trial was begun in the startbox, with all doors open. Following this trial *S* was placed in the detention cage for 5 min., after which it was returned to its home cage and fed for 1 hr. On the second day of accommodation the essential procedure of the first day was repeated, except for the initial 15-min. period and except that two trials were run from each starting position (to a total of 14 trials).

Training.—The *S*s were started in squads of four. They were given 12 training trials per day for 6 days. Each endbox was used four times, on a random schedule, in every set of 12 trials. Two pellets (weighing approximately .10 gm.) were in the foodcup in one of the endboxes only. The *S* was placed in the detention cage for 20 sec. between trials.

On the first training day an initial trial from the startbox was given with the neutral endbox in place and with all doors open, prior to the regular 12 trials with the startdoor closed. On each trial *S* was placed in the startbox and the startdoor opened 3 to 5 sec. thereafter. If *S* did not complete the run to the endbox within 180 sec., it was gently guided down the runway to the endbox floor switch. It was allowed 10 sec. in the endbox; on rewarded trials, if it had not eaten in that time, it was placed by the foodcup and given an additional 10 sec. to eat. At the end of the 12 trials *S* was placed in the individual detention cage for 5 min., after which it was returned to the home cage. When all four *S*s in a squad were back in the home cage, they were fed for 1 hr.

Extinction.—The *S*s were placed in the four groups randomly, within squads. Massed extinction trials were given, at an interval of 20 sec., to a criterion of two successive failures to enter the endbox with a minimum of 24 trials. Of the nine *S*s in each group, three had been run to each type of endbox under reward conditions. Group P now found only the formerly positive endbox, Group N only one of the formerly negative endboxes, Group NN both of the formerly negative endboxes (in the same order in which they had occurred during training), and Group NNP all three endboxes (in the same order in which they had occurred during training).

RESULTS

Training

By the sixth day of training essentially the same nearly asymptotic starting and running times had been reached by all four groups. In starting time, for example, all of the group medians for the 12 trials were within the range of 1.0 to 1.1 sec. In running time, three of the medians were 1.6 and one was 1.7 sec. No difference among groups was apparent in rate of approach to these values.

Time required to traverse the 18 in. between floor switches in the endbox is a somewhat more interesting measure because it provides a direct index of *S*'s discrimination between positive and negative goal conditions. Here again the groups reached essentially similar and nearly asymptotic running times. A clear and increasing difference between goal times in the positive and negative endboxes occurred from the second day on. With all 36 animals considered, median daily times were as follows: for the positive endbox, 7.3, 4.0, 2.3, 1.6, 1.5, and 1.3 sec; for the two negative endboxes, 8.0, 5.5, 5.0, 4.6, 4.8, and 4.9 sec. On the last day of training, each of the 36 *S*s ran faster in the positive endbox than in the negative endboxes.

Extinction

Trials to extinction.—Increasing resistance to extinction was found in the direction predicted. Trials to an extinction criterion of two successive failures to complete the

TABLE 1

Mean Number of Trials to Reach Goal Time Criterion 1 (First 10-sec. Score) and Goal Time Criterion 2 (Two Successive 10-sec. Scores, with Criterion 1 Scores Excluded)

Criterion	Group			
	N	P	NN	NNP
1	3.1	10.7	4.0	4.8
2	8.0	3.9	4.1	15.6

running response within 90 sec. gave the following mean values: Group P, 45.7; Group N, 50.7; Group NN, 60.3; and Group NNP, 94.3. This difference among groups was highly reliable ($F = 12.11$, $P < .001$ for *df* 3/32). Direct comparisons of pairs of groups indicate that only the difference between Group NNP and the other groups was reliable (mean difference of 20.5 required for $P = .05$).

Trials to a less stringent criterion of extinction, the first failure to respond, gave essentially the same group differences but at a somewhat lower level of reliability ($F = 4.53$, $P < .01$, for *df* 3/32).

Group comparisons were also made for extinction of the running response within the goalbox. Two criteria were used: (*a*) number of trials to the first 10-sec. goal time; and (*b*) number of trials to two successive 10-sec. goal times. Here the influence of the formerly positive endbox is more clearly shown, as would be expected on the basis of the discrimination evident in training. The mean numbers of trials to meet each criterion are shown for the four groups in Table 1. Group P required more trials than Group N ($P < .05$) and than Groups NN and NNP ($P < .10$) for Criterion 1 and Group NNP reliably more than all other groups ($P < .05$) for Criterion 2.

A direct comparison of positive endbox times with negative endbox times within Group NNP gave the following results: 8.4 mean trials to the first 10-sec. trial for the positive, 3.4 for the negative; 10.3 additional mean trials to two successive 10-sec. trials for the positive, 8.8 additional for the negative.

Starting and running times.—The increments in starting and running times showed the same intergroup relations as were found for the trials-to-extinction measure. Interpretation of these gross data is somewhat difficult, however, because of the random admixture of endbox conditions in Group NN and Group NNP.

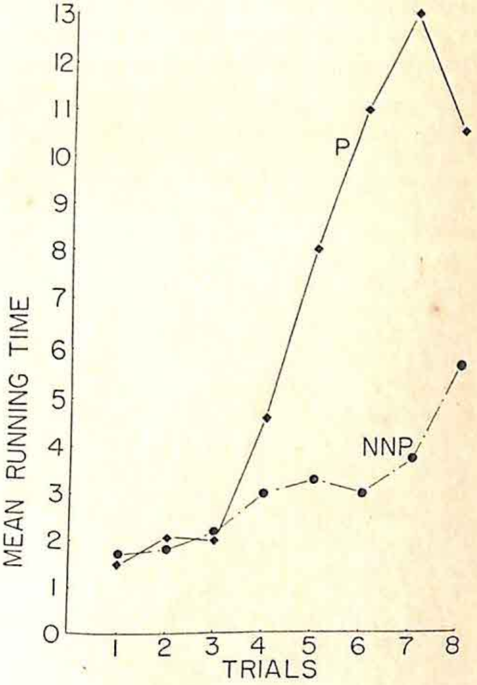


FIG. 1. Mean running times following presentation of previously positive endbox during extinction for Group NNP and comparable trials for Group P.

TABLE 2

ANALYSIS OF VARIANCE ON RUNNING TIME FOLLOWING PRESENTATION OF PREVIOUSLY POSITIVE ENDBOX DURING EXTINCTION FOR GROUP NNP AND COMPARABLE TRIALS FOR GROUP P[a]

Source	*df*	*MS*	*F*
Groups	1	399.03	8.82*
Error	14	48.44	
Trials	7	127.19	6.65**
G × T	7	56.18	2.94**
Error	98	19.12	

* $P < .05$.
** $P < .01$.
[a] One *S* in Group P had a criterion trial too early to be used in this analysis; to keep *N*s equal, one *S* in Group NNP was randomly discarded. If these *S*s were used, giving the Group P *S* an arbitrary score, differences would be magnified.

For this reason more analytic running-time comparisons are reported.

The effect of interpolation of negative endboxes during extinction may be seen in a comparison of the mean running time on the trial *immediately following* entrance into a formerly positive endbox (Group NNP) with that on comparable trials where only the previously positive endbox is presented (Group P). These data are shown graphically in Fig. 1. The clearly superior maintenance of running speed in Group NNP was highly reliable, as is shown in the statistical analysis presented in Table 2.

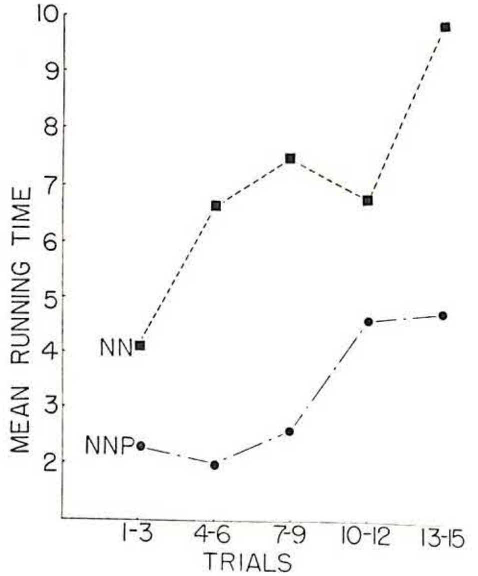


FIG. 2 Mean running times following presentation of previously negative endboxes during extinction for Group NNP and comparable trials for Group NN.

TABLE 3

ANALYSIS OF VARIANCE ON RUNNING TIME FOLLOWING PRESENTATION OF PREVIOUSLY NEGATIVE ENDBOXES DURING EXTINCTION FOR GROUP NNP AND COMPARABLE TRIALS FOR GROUP NN

Source	*df*	*MS*	*F*
Groups	1	936.32	4.806*
Error	16	194.81	
Trials	14	80.73	<1
G × T	14	64.23	<1
Error	224	98.09	

* $P < .05$.

The effect of interpolation of a positive endbox may be seen in a comparison of the mean running time on the trial immediately following entrance into a formerly negative endbox (Group NNP) with that on comparable trials where only the two formerly negative endboxes are presented (Group NN). These data are shown in Fig. 2, for pooled blocks of three trials, in order to show the trends more clearly. Here the group difference, but not the Group × Trials interaction, was reliable, as shown in Table 3. The mean running times were 3.25 sec. for Group NNP and 6.98 sec. for Group NN.

No reliable differences were found in running time between comparable trials in Groups N and NN ($F < 1.00$). No important effect can be attributed to the interpolation of a single formerly negative endbox in the absence of any positive endboxes.

DISCUSSION

The results of the experiment clearly support the prediction that resistance to extinction is a function of the degree of reproduction of the training conditions.

These data are consistent with the discrimination hypothesis (Bitterman, Fedderson, & Tyler, 1953; Elam, Tyler, & Bitterman, 1954). However, the failure of the simple N condition to show reliably greater resistance to extinction than the simple P condition does not support this interpretation. Also, the discrimination hypothesis was contraindicated by the results of an experiment previously reported (Marx, 1958). A recent experiment by Brown and Bass (1958) may be cited as failing to substantiate a discrimination or perceptual-change interpretation of extinction, while pointing to the importance of the variety of stimulus conditions as a strong contributor to resistance to extinction.

The most interesting aspect of the present data concerns the analytic running-time measures shown graphically in the figures. The main conclusion to be drawn from these comparisons is that the positive endbox, when interpolated among the negative endboxes on successive trials, is the critical factor in the situation. The statistically reliable differences found when Group NNP is compared with Group P and Group NN are to be contrasted with the unreliable difference found in the comparison of Groups N and NN. The importance of the interpolation of negative cues in extinction is also indicated by the comparison of goal time extinction scores for Groups P and NNP (see Table 1); Group P ran longer to the first, but Group NNP longer to the second consecutive, failure to enter the endbox.

According to the present interpretation the formerly positive endbox acts as a strong motivator when interpolated among formerly negative endboxes. Apparently the contrast serves to maintain *S*'s instrumental motivation in the same way as the cup–no cup condition did in the previous experiment.

Theoretical interest attaches to the fact that the differential effect was obtained by "interpolation" of the formerly positive endbox, as compared with a formerly negative endbox, when running speed following presentation of comparable negative endboxes was measured (Fig. 2). This fact suggests that a simple interpretation of these over-all differences solely in terms of the differential extinction of the secondary reinforcing properties of a positive cue is inadequate. The effect appears to be a more general motivational one, since it apparently carries over to running responses following negative cues, although the interpolated presentation of the positive cue as the focal point of the effect confirms the results of the previous experiment (Marx, 1958).

SUMMARY

This experiment tested the hypothesis that resistance to extinction of an instrumental response is a function of the degree of reproduction of training conditions. Four groups of 9 rats each were given three distinctive endbox conditions during training, but food was found in only one of these. In extinction, one group was run with only the formerly positive endbox, a second with only one of the formerly negative endboxes, a third with both negative endboxes, and a fourth with both negative and the positive. The results were in accordance with the prediction, with resistance to extinction among groups increasing in the order mentioned above. More analytic comparisons of running times showed the formerly positive cue when interpolated among formerly negative cues to be the major factor in the maintenance of the running response. The results were interpreted as support for the general proposition that resistance to extinction is a function of the maintenance of instrumental motivation.

REFERENCES

BITTERMAN, M. E., FEDDERSON, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

Brown, J. S., & Bass, B. The acquisition and extinction of an instrumental response under constant and variable stimulus conditions. *J. comp. physiol. Psychol.*, 1958, **51**, 499–504

Elam, C. B., Tyler, D. W., & Bitterman, M. E. A further study of secondary reinforcement and the discrimination hypothesis. *J. comp. physiol. Psychol.*, 1954, **47**, 381–384.

Marx, M. H. Resistance to extinction as a function of continuous or intermittent presentation of a training cue. *J. exp. Psychol.*, 1958, **56**, 251–255.

(Received June 22, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

REPLICATION REPORT: AN ATTEMPT TO OBTAIN INHIBITION WITH REINFORCEMENT[1]

STEPHANIE S. FUCHS[2]

Reed College

Kendrick (1958) has reported the experimental validation of a deduction from Hullian theory which predicts the extinction of a continuously reinforced response. The present study is a replication of his experiment with respect to the pre-experimental handling of *S*s, the 30 trials per day, the criterion of extinction, the 20-hr. thirst drive, the $\frac{1}{4}$-cc. water reward, and the other details described in his article (Kendrick, 1958, p. 314) except as follows: Kendrick's *S*s were chosen from populations characterized by extremes of emotionality; *S*s in the present study were six male Sprague-Dawley albino rats, 118 days old at the beginning of the experiment. The *S*s were run one at a time, in the middle of a 9-hr. dark period; Kendrick's *S*s were run two at a time, after the first 6 hr. of a $17\frac{1}{2}$-hr. light period.[3] As in the Kendrick study, Days 1–3 were training days during which *S*s explored the alley; Day 4 was the first day of testing.

No extinction of the running response occurred in the present study within 41 days of 30 trials per day. In comparison, the mean number of days to extinction for Kendrick's *S*s was 32 days, and his longest-running *S*s extinguished on Day 41. Thus the Kendrick results were not confirmed.

Figure 1, showing median running times for the group for the testing days, indicates that running time was at asymptote for over 25 days. Latency results were very similar and less variable from day to day. Figure 1 also indicates that running times were very consistent on the first few trials of each day throughout the testing series, while, in contrast to this, on the last six trials, running times were extremely variable and much longer. Figure 2, showing intraday variation in running time, indicates that running times for the last six trials were consistently longer than those for the first six.

The behavior of *S*s was recorded for all trials; analysis of these data for those trials which were longer than 4 sec. indicates that *S*s are more active during the long later trials than during the long trials at the beginning of a day. During the early long trials, *S*s tended to walk slowly down the alley, sniffing

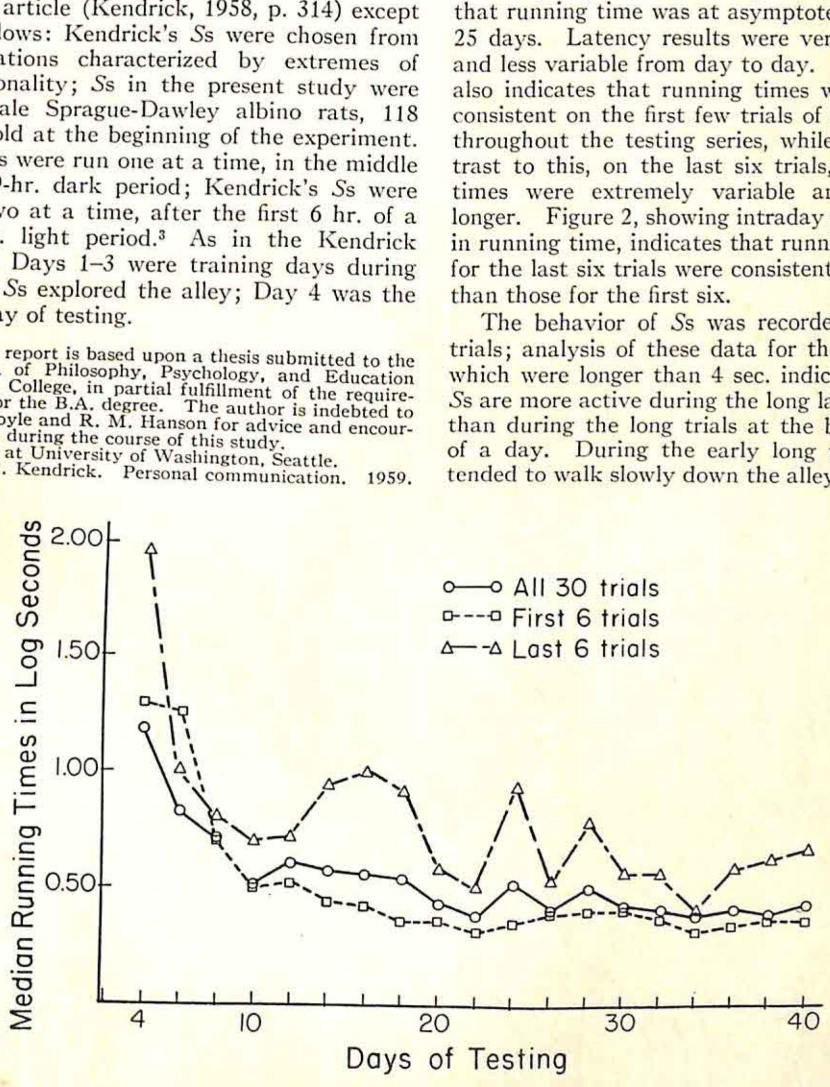


FIG. 1. Performance curves based on the median of the six median running times for all trials, for the first six trials, and for the last six trials per day.

---

[1] This report is based upon a thesis submitted to the Division of Philosophy, Psychology, and Education of Reed College, in partial fulfillment of the requirements for the B.A. degree. The author is indebted to R. E. Boyle and R. M. Hanson for advice and encouragement during the course of this study.

[2] Now at University of Washington, Seattle.

[3] D. C. Kendrick. Personal communication. 1959.

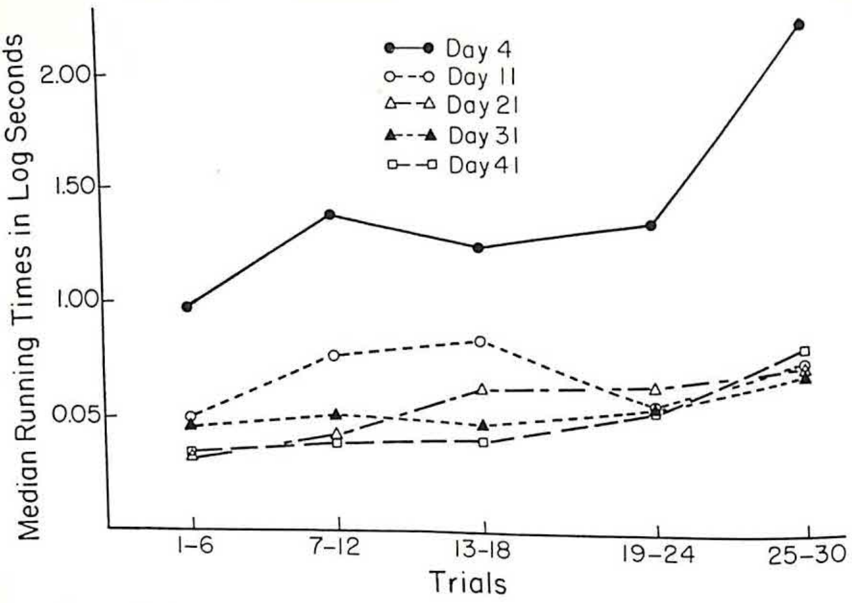


FIG. 2. Performance curves based on the median running times for successive fifths of the day's testing on selected days.

its floor and sides. This behavior rarely occurred at the end of a day's testing and is in marked contrast to the typical behavior during long late trials, which included such activity as running on the edge of the alley and retracing several times in the alley; these latter types of behavior almost never occurred in the early long trials of a day.

The increased activity during the long later trials of a day when *S*s are presumably more fatigued precludes an explanation of long running times in terms of fatigue of the running response. No evidence was found for the conditioned inhibition of the running behavior, since almost no variability or increase appeared in the latencies and running times on the first trials of a day, when *S*s were least fatigued.

It is possible that extinction would have occurred had *S*s been run for some longer period. Factors which might affect the length of time necessary to achieve extinction are the strain of *S*s, the amount of emotionality, and possible differences in activity level due to the differences in the relationship of the time of testing to the light/dark cycle. However, in the present study, the analysis of *S*s' behavior indicates that increased variability in running times is due to increased activity, rather than to increased fatigue.

REFERENCE

KENDRICK, D. C. Inhibition with reinforcement (conditioned inhibition). *J. exp. Psychol.*, 1958, **56**, 313–318.

(Received November 16, 1959)

Journal of
Experimental Psychology

VOL. 59, NO. 6 JUNE 1960

EFFECTS OF INCREMENTS OF REINFORCEMENT IN HUMAN PROBABILITY LEARNING[1]

MAYNARD W. SHELLY II

Office of Naval Research

There are indications in the literature that even though relative frequencies of reinforcement may have proportionate effects on response frequencies asympototically, they may not have such proportionate effects at preasymptotic levels (e.g., Grant, Hake, & Hornseth, 1951). Such indications can be observed when trial or block number replaces reinforcement conditions as the parameter of the family of performance curves, and reinforcement conditions replace trial or block number on the abscissa in the usual graphs of results. From similar plots there are also faint evidences (e.g., Neimark, 1956) that any lack of preasymptotic proportionality may be affected by the type of reinforcement, i.e., whether it is contingent or noncontingent. This experiment was conducted to determine how the relative frequency of a response with some fixed probability of reinforcement is affected by variations in the probability of reinforcement of alternative responses.

To study the effects of variations in the probability of making alternative responses, this investigation employed a situation approximating the usual probability learning conditions. Three alternatives were used, and *S*s were asked to predict which of the three alternatives would be correct on a given trial. None, one, two, or three alternatives could be correct on any one trial. Because of the indications that contingent and noncontingent reinforcement might produce different effects on another alternative, even for a constant probability of reinforcement, both conditions of reinforcement were used.

This study uses three, rather than two, response alternatives for two principal reasons. The first is that with three alternatives it is possible to have more sizable variations in the

---

[1] This research was conducted while the author was at the University of Wisconsin. He wishes to express his appreciation to the Graduate School of the University of Wisconsin for providing funds for the experimental apparatus employed, and to the Social Science Research Council for providing the author with a Research Training Fellowship for the period during which the experiment was completed. He also wishes to thank the Office of Naval Research for providing the author with time, and the Department of Psychology, Northwestern University for the facilties to complete the final drafts of this paper.

sum of the probabilities of reinforcement on two of the three alternatives and still have the probability of reinforcement on the third a maximum. Secondly, it is only when three or more alternatives are used that contingent and noncontingent reinforcement are fully differentiated.

METHOD

Subjects.—The *S*s were 170 students (approximately evenly divided between women and men) from elementary psychology courses at the University of Wisconsin.

Apparatus.—The apparatus consisted of a booth with a flat black interior in which *S* sat, and equipment outside the booth which regulated the experimental conditions. In the booth an aluminum chassis box (8 in. × 17 in.) was mounted on a sloping table. It contained 16 four-pole snap switches, above each of which was mounted a 6-v. green-domed lamp. For the conditions of this study, only the three switches and their associated lamps were uncovered. A yellow-domed 6-v. lamp was also mounted in front of *S*. This informed him that a new trial had begun. When it went off, *S* was to make a choice of one of the three switches before him.

The equipment outside of the booth consisted largely of a timing motor, Western Union tape transmitters, and lamps on a panel which permitted *E* to record *S*'s responses. There were also a number of rotary switches which permitted *E* to change the correspondence between channels of the Western Union tape transmitters and *S*'s response alternatives.

Design.—The experiment consisted of 17 conditions, 10 different *S*s serving in each condition. In all 17 conditions at least one alternative would lead to a correct prediction on 75% of the trials if always chosen. This alternative was combined with two other response alternatives. The reinforcement probabilities on these other two alternatives were combinations of 0, .25, .50, and .75. Random binary sequences were established independently for each alternative, so that even when the probabilities summed to one, there were trials on which no switch was correct and trials on which one or more switches would be correct.

The two types of reinforcement were contingent and noncontingent reinforcement. When *S* served under contingent reinforcement, he was only informed as to whether or not he had made a correct prediction. If he had been correct, the lamp above the alternative he had chosen came on. If he had been incorrect in his prediction, no lamp came on. When noncontingent reinforcement was used, a lamp came on above each correct alternative (if there were any) irrespective of whether *S* had made a correct or incorrect prediction.

TABLE 1

COMBINATIONS OF REINFORCEMENT PERCENTAGES USED IN THE EXPERIMENTAL CONDITIONS

75-0-0[a]		
75-0-25	75-25-25	
75-0-50	75-25-50	75-50-50
75-0-75	75-25-75	75-50-75

Note.—The numbers in the table represent the percentages of times the first, second, and third alternatives, respectively, would be reinforced if constantly chosen.

[a] This condition was not completed under noncontingent reinforcement because *S*s were responding perfectly after only a few trials, and a factor of boredom would have,been too strong a confounding influence.

The design is presented in Table 1. Here the first "75" refers to the presence of an alternative which was reinforced 75% of the time. The two numbers following the initial "75" give the percentages of reinforcements which would be received were either ot the other alternatives constantly chosen. The effects of any consistent position preferences among *S*s were eliminated by permuting the positions of the various alternatives among *S*s.

Procedure.—The *S*s in this experiment served individually. When they reported to the laboratory, they were assigned to one of the 17 conditions on a random basis with the restriction that no more than 10 *S*s be assigned to any one condition. Upon being seated in the experimental booth, all *S*s were given the same standard set of instructions (except for modifications appropriate to the conditions of reinforcement). These instructions told *S*s that the experiment was designed to study their abilities to predict future events, and that the object of the experiment was to make as many correct predictions as possible. Also included were instructions regarding the mechanics of the experiment. It was strongly suggested that they were to make one choice on each trial, and never more than one. A brief demonstration was then given.

Each *S* then made a total of 320 predictions. Each trial was of 4-sec. duration.

RESULTS

In Fig. 1 are presented the performance curves for the 75% alternative for the conditions of the present experiment. At the termination of each curve are two numbers separated by a hyphen. These numbers represent the percentages of times that the two alternatives not plotted for that group could lead to reinforcement in 100 trials. The number in parentheses associated with each curve refers to the asymptote which would be approached were *S* responding proportionately (in the following sense) to the percentage of reinforcement. A response will be considered to be proportionate if the proportion of times an alternative is chosen is equal to the ratio of the number of times that alternative can be reinforced to the sum of the number of times all alternatives can be reinforced. For the 75-25-25 condition the ratio for the 75% alternative would be $75/(75 + 25 + 25)$. As can be seen, the departures from such proportionality are greatest when the denominator of the above ratio is smallest.

In the introduction it was stated that there are suggestions that increments in amount of reinforcement assigned to other alternatives do not always have proportionate decremental effects on the response alternative being analyzed. Such a suggestion seems to be further confirmed by the performance curves in Fig. 1. The *D*s in Fig. 1 are the distances between curves *when* the totals of the two numbers identifying the curves involved differ by 25. (To keep the graphs relatively simple only two of the possible *D*s are shown.) To consider an example, take Part a of Fig. 1; here $25 + 0$ and $50 + 0$ differ by 25 and an arrow is drawn between the two curves (for the last block of trials only) and labeled by *D*. It can be observed that the *D* distances are not always proportionate to the differences between the numbers included in the parentheses. In an attempt to analyze such relationships more formally, the following type of decision procedure was used.

We can specify the percentages of reinforcement for two alternatives and let the third vary over the values 0, 25, 50, and 75, which stand for percentages of reinforcement. Denote the third alternative, which can vary, by X. Then, using { } to denote a set, {75-0-X} is, as an example, the set of conditions {75-0-0, 75-0-25, 75-0-50, 75-0-75}. Such sets of conditions will be referred to as *C sets*. It will be observed that when the members of a C set are written in the above ordering, the percentage of reinforcement for the last alternative increases by steps of 25; i.e., the X values increase by steps of 25. By using the *D* distances as ordinate values, and X values on the abscissa for a fixed C set of conditions we get a family of curves, the parameter of which is trial block number. For each of the two types of reinforcement there are four C sets: {75-0-X}, {75-25-S}, {75-50-X}, and {75-75-X}. Clearly, the conditions of the experiment usually do not belong to only one C set.

Because X can assume four values, there are three *D* distances possible, and hence each curve, for a given block of trials, based upon a given C set, contains three points, the first of which is at an X value of 25. Each of these three points can be interpreted as the decrease in number of times the first alternative (the alternative which in every condition could lead to a reinforcement 75% of the time) is responded to as a

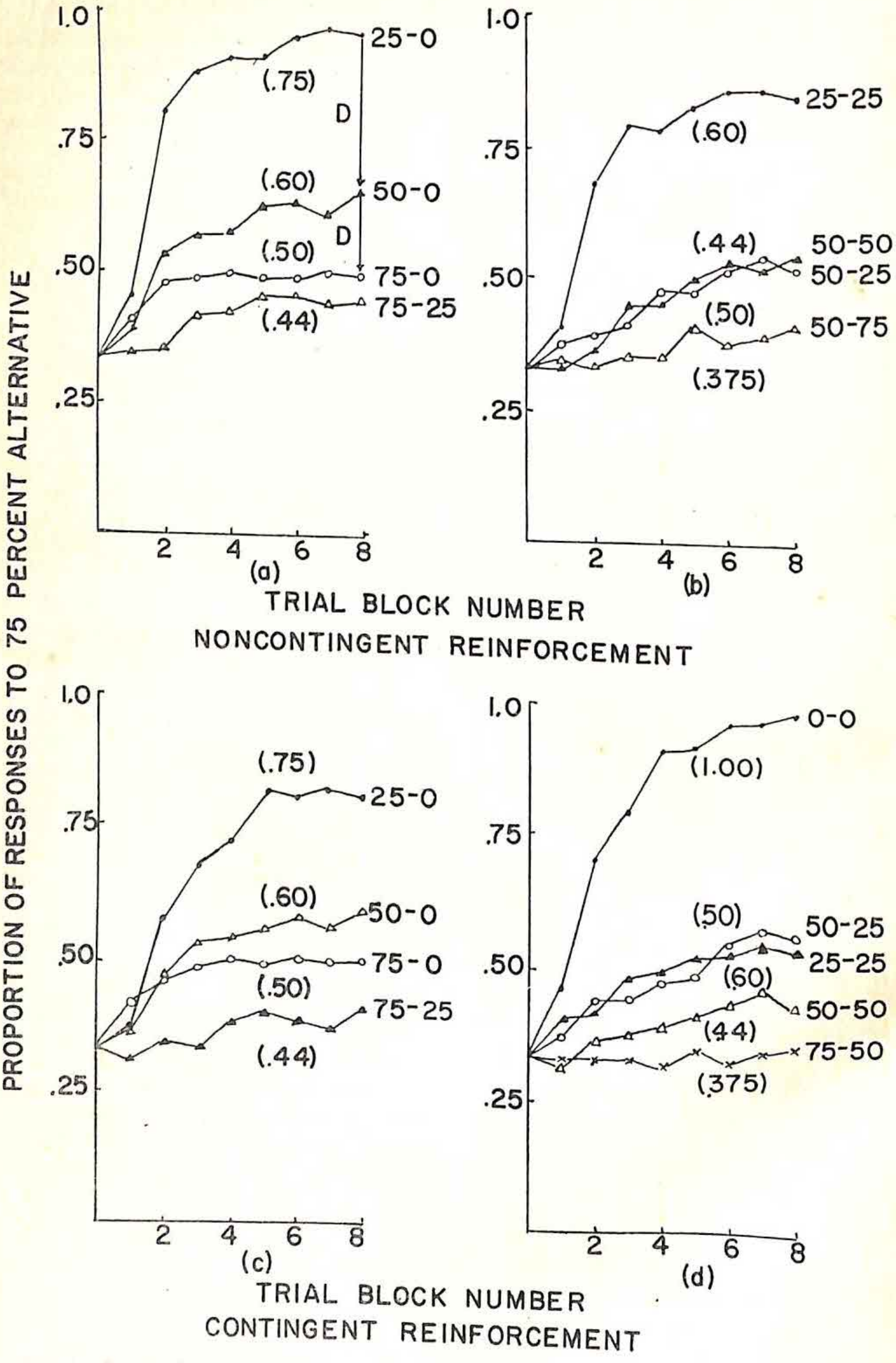


FIG. 1. Proportion of responses to the 75% alternative in all conditions for Trial Blocks 2, 4, 6, and 8. The division of performance curves between Parts a and b, and between Parts c and d is simply to make them more easily readable.

result of an increase in the X value by 25, e.g., from 0 to 25, or from 25 to 50. It can easily be computed, using the definition of proportionality in this paper, that the maximum point on each of these curves should be the first point, i.e., the D distances between 75-0-0 and 75-0-25 for example.

Families of curves based upon the C sets were plotted for the second, fourth, sixth, and eighth block of 40 trials. Sequences of one or two hypotheses were then tested. The first was that twice the maximum point on each curve was greater than the sum of the other two. This simply tested the statistical hypothesis that there was a statistically significant maximum point. The procedures used in this and the following hypothesis are those outlined in Grant (1956). For example, if the second point were the greatest, the hypothesis tested would be:

$$-(\mu_0 - \mu_{25}) + 2(\mu_{25} - \mu_{50}) - (\mu_{50} - \mu_{75}) = 0 \quad [1]$$

where μ_0, for example, is the mean for that condition of the C set for which $X = 0$.

If this hypothesis were not rejected, then the hypothesis that the sum of the two greatest points was greater than twice the remaining point would be tested. For example, if this happened to be the first two versus the third point, the hypothesis tested would be:

$$(\mu_0 - \mu_{25}) + (\mu_{25} - \mu_{50}) - 2(\mu_{50} - \mu_{75}) = 0 \quad [2]$$

where μ_0, μ_{25}, etc. have the same meanings as in [1]. If either one of the hypotheses were rejected, then a statistical maximum would be reached at one, or at two points. These inferred statistical maxima, max(D), are given in Table 2. It is very evident that max(D), when it occurs, does not always occur at an X value of 25. Indeed, 62.5% of them occur at an X value of 50.

TABLE 2

VALUES OF X FOR WHICH MAX(D) OCCURS

C set	Trial Block Number			
	2	4	6	8
Noncontingent Conditions				
{75-0-X}[a]	50	50	50	50
{75-25-X}	50	50	50	50
{75-50-X}	25	—	—	(25, 75)
{75-75-X}[b]	—	—	—	—
Contingent Conditions				
{75-0-X}	—	(25, 50)	—	50
{75-25-X}	—	—	(25, 75)	(25, 75)
{75-50-X}	—	—	—	—
{75-75-X}[b]	25	—	—	—

[a] For 75-0-0 the value of 1.00 was used for the proportion of responses to the 75% alternative and the statistical procedures adjusted accordingly.

[b] For 75-75-75 the value of .333 was used for the proportion of responses to the 75% alternative and the statistical procedures adjusted accordingly for use of a nonempirical value.

DISCUSSION

Obviously the more complete implications of this study can only be arrived at through an extended mathematical analysis. There are, however, several things about the findings which can be emphasized. First, the max(D) values do not usually occur at an X value of 25. This implies that the probabilities of reinforcement are not being responded to proportionately. This result agrees with the findings of Gardner (1957) and Cotton and Rechtschaffen (1958). Secondly, there is considerable consistency in the X values at which the max(D)s occur for different blocks of trials. This implies some consistency in the relative learning rates. Thirdly, 62.5% of the max(D)s occurred at an X value of 50. That is, the greatest decrease in responding to the 75% alternative occurred when the probability of reinforcement on the variable alternative in a C set reached maximal uncertainty, i.e., changed from a probability of .25 to a probability of .5. It is possibly this last finding which is most suggestive for future research.

SUMMARY

An experiment was conducted to determine how the relative frequency of a response with some fixed probability of reinforcement is affected by variations in the probability of reinforcement of alternative responses. The experimental situation approximated the usual probability learning conditions. The *S*s were asked to predict which of three events would occur on any one trial. One of the alternative responses had a probability of reinforcement of .75 for all *S*s. The second and third responses had probabilities of reinforcement varying from 0 to .75 in steps of .25 for the various experimental groups. The results were found not to be in agreement with a prediction based upon a generalization of the meaning of proportionality for two alternative responses. Instead it was found that, when the reinforcement value of the first two alternative responses was kept constant, the greatest decrement in responding to the first (the .75 alternative) occurred when the probability of reinforcement on the third alternative increased from .25 to .5.

REFERENCES

COTTON, J. W., & RECHTSCHAFFEN, A. Replication report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

GARDNER, R. A. Probability learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

GRANT, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

NEIMARK, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

(Received June 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

IDENTIFICATION OF VISUAL CORRELATIONAL SCATTERPLOTS[1]

IRWIN POLLACK

Operational Applications Laboratory, Air Force Cambridge Research Center

A useful tool for the examination of interaural phenomena is the description of auditory signals at the two ears in terms of their correlation with each other (Pollack & Trittipoe, 1959; Zerlin, 1959). The aim of the present paper is to describe initial studies with a visual correlational display. This display, we believe, provides a useful tool for the examination of dynamic shape perception and the behavioral testing of statistical hypotheses.

The student of psychological statistics is familiar with the correlational scatterplot. If two variables, *A* and *B*, strongly share a single common factor, *C*, the correlational scatterplot of *A* versus *B* will be confined to a small area about the diagonal. If the two variables weakly share the common factor, *C*, the correlational scatterplot may be diffuse and may cover a large extent of the entire *A* versus *B* area. In terms of components of variance, the correlation (squared) is equal to the ratio of the variance common to each of two variables to the total variance (common variance plus independent variance associated with each of the two variables).

Licklider and Dzendolet (1948, Fig. 3, p. 123) have demonstrated a simple procedure for producing correlational distributions. Three independent random noise generators, *A*, *B*, and *C*, are employed. If the output of Noise Generator *A* is sent to the x coordinate of an oscilloscope and the output of Independent Noise Generator *B* is sent to the y coordinate, a diffuse correlational plot with a circular envelope will be obtained. Since *A* and *B* are independent random distributions, a correlation of zero will be obtained. If, on the other hand, the output of Noise Generator *C* alone is sent to both coordinates of the oscilloscope, a thin line at 45° will be traced if the x and y coordinate gains are equal. Since a single distribution is sent to both coordinates, a correlation of 1.0 will be obtained.

If *C* is mixed with *A* and sent to the x coordinate; and if *C* is mixed with *B* and sent to the y coordinate, intermediate correlations can also be obtained. The greater the relative contribution of the common source, *C*, the higher will be the correlation. The greater the relative contribution of the independent sources, the smaller will be the correlation. And, in general, the correlation (squared) will be equal to the ratio of the common noise power, *C*, over the total noise power $(C + A)$ or $(C + B)$. Licklider (1948, Fig. 4, p. 155) presents oscillographic pictures traced out by a rapidly moving electron stream for a wide range of correlations. Licklider and Dzendolet (1948, Fig. 2, p. 122) also present oscillographic pictures obtained by discrete sampling of correlational distributions.[2]

---

[1] This report is AFCRC TR 59-14 of the Air Force Cambridge Research Center. This research supports Project 7682, Information Processing by the Human Operator, of the Air Research and Development Command Program in Human Performance. The writer wishes to acknowledge the able assistance of William Trittipoe in all phases of the testing program, and John Schjelderup with respect to the experimental equipment.

METHOD

The task of *S* was to identify one of two correlational distributions presented on an oscilloscope. In terms of three independent noise generators, a reference correlation was obtained by routing $(C + A)$ to one coordinate of the oscilloscope and by routing $(C + B)$ to the other coordinate. In all tests, $A = B$; and $(C + A) = (C + B)$. A second, higher, correlation was obtained by mixing a higher level of the common source, C', with the two independent noise sources. The task of *S* was to identify whether the correlation was produced by $(C + A)$ and $(C + B)$ or by $(C' + A)$ and $(C' + B)$. In terms of the verbalization of *S*'s task, he was instructed to press his voting lever "down" for the "stubbier" of the two scatterplots and to press his voting lever "up" for the "thinner" of the two scatterplots. It is important to note that the power sum of $(C' + A)$ was equated to $(C + A)$ in order that distributional differences, rather than differences in amplitude, would serve as the relevant cue for discrimination.

Factors which facilitated rapid learning of the task were: (*a*) immediate confirmation of correctness of response was provided; (*b*) a consistent direction of movement of *S*'s voting lever was assigned to the "thinner" of the two correlations; (*c*) *S*s worked in a group and the score of each *S* was announced to the entire group after each set of 30 or 60 presentations.

The oscilloscopes employed were Dumont Model 304-A, one with a fast blue P-11 phosphor, one with a medium persistence green P-1 phosphor. Since no consistent difference was observed between the results yielded by the two tubes, only the average results are reported. Approximately three-quarters of the 5-in. diameter was employed for the display. The noise generators were Grason-Stadler Model 455 B, set at 20 kcps. The noise spectrum was uniform between 100 and 6800 cps after being passed through a band-pass filter with an attenuation characteristic of 18 db/octave. Low-pass filtering was obtained by sharp *m*-derived filters, Gertsch Model BP-1. A variable phase-shift circuit was employed in order to sharpen the scope image for high noise frequencies. The interval between successive correlation displays was 4 sec.

Discrete samplings from the correlational distributions were obtained in the following manner. The two correlational distributions were first established, and the intensity control of the oscilloscope was reduced so that a picture was not visible. The intensity axis, or *z* axis, of the oscilloscope was then pulsed to yield short discrete samples from the correlational distribution. The pulse duration was 6 μsec., obtained from a Measurements Corporation 79-B pulse generator.

In all, eight separate experimental series (to be described separately) were carried out. In each series, a reference correlation $(C + A)$ was established, and a wide range of variable correlations $(C' + A)$ was explored in order that the point of 75% correct two-choice identification could be established. In all tests except Series IV, the average seating distance from the scopes was about 18 in. In Series IV, the distance was 54 in. A nonpulsed display was employed in Series I–V; a sampled discrete-pulsed display was employed in Series VI–VIII.

Series I consisted of four *S*s, all laboratory personnel experienced in psychophysical testing. Series II–V employed six *S*s, five female college undergraduates and one of the *E*s. Prior to the testing series, the undergraduates had not participated in psychophysical studies. Series VI–VIII employed six *S*s, four female college undergraduates and both *E*s, all experienced in psychophysical studies. Each presented datum point for for Series II–VIII represents the interpolation from a psychophysical function based on about 1800 observations—6 *S*s, 4–6 variable correlations, and 60 observations by each *S* for each variable correlation, except Series V the corresponding number of observations was about 5600.

RESULTS

Series I: Preliminary Tests

The purpose of Series I was to compare the results of visual tests with the results of auditory tests that were carried out with the same noise distributions (Pollack & Trittipoe,

---

[2] Even the excellent still photographs in the two references fail to convey the dazzling dancing parade of correlation patterns available for study. To the serious reader interested in dynamic displays with information of well defined statistical properties, a lash-up of the noise generators and oscilloscope is recommended to taste the full richness and flavor of the available materials.

1959). In the auditory tests, $(C + A)$ or $(C' + A)$ was presented to one earphone and $(C + B)$ or $(C' + B)$ was presented to the other. The abscissa of Fig. 1 is the reference correlation (squared); the ordinate is the difference between the reference correlation (squared) and the variable correlation (squared) required for 60, 75, and 90% correct identification. The visual tests are represented by solid lines. The thin dashed curve represents 60% correct identification for the auditory tests.

Two main points are suggested by Fig. 1: (*a*) In light of the vastly different tasks with the auditory and visual displays, the form of the correlation identification function for auditory and visual tests is strikingly similar. The 60% correct auditory contour nearly overlaps with the 90% correct visual contour. (*b*) The three visual correlation identification functions exhibit roughly the same general form. Because the two-choice identification function is nearly linear about its midpoint, the 75% correct identification level was selected to summarize the results of all of the other tests.

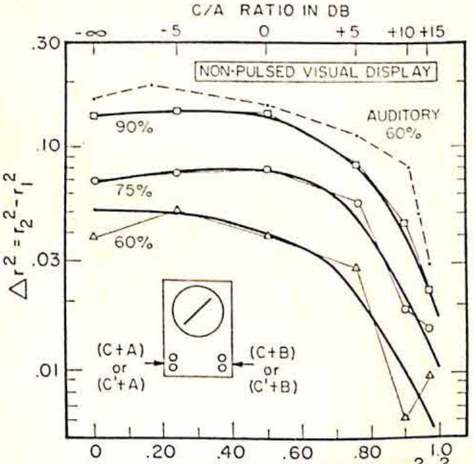


FIG. 1. Comparison of auditory and visual correlational displays. The abscissa is the reference correlation (squared). The ordinate is the difference between the required correlation (squared) for 60, 75, and 90% correct identification and reference correlation (squared). The ordinate is scaled logarithmically. The dashed line represents the results of auditory tests for 60% correct identification. Tests of Series I.

Series II: Effect of Duration

The results of Series II are presented in Table 1. A given testing condition is defined by the display duration (across the top) and the reference correlation (along the left side). For each condition, three entries are presented: (*a*) the C'/A ratio in decibels required for 75% correct identification when paired with the reference correlation on the left; (*b*) the equivalent correlation (squared); and (*c*) the equivalent Fisher z transformation score. The three corresponding entries are also presented for the reference conditions on the left.

The effect of duration is not marked except for the conditions of high reference correlation where extremely large differences were needed to obtain 75% correct identification. (The 75% correct identification level could not be obtained for the 10-msec. display for the two highest reference correlations within the ranges of conditions available.)

The ratio between C and C' (in terms of the decibel difference required for 75% correct identification) furnishes one measure of relative sensitivity. In terms of this measure, maximum sensitivity is obtained near a reference correlation (squared) of about 0.5 with reduced sensitivity for lower and higher reference correlations.

Another measure of sensitivity is the difference between the correlations (squared) required for 75% correct identification. The results

TABLE 1

IDENTIFICATION OF TWO VISUAL CORRELATIONS AS A FUNCTION OF THE REFERENCE CORRELATION AND DISPLAY DURATION

(For each condition, three entries are presented: (*a*) the C'/A ratio in decibels which was correctly identified 75% of the trials from the reference C/A ratio listed on the left; (*b*) the equivalent correlation coefficient (squared); and, (*c*) the equivalent Fisher z-score transformation.)

Reference Correlation	Measure	Duration in Milliseconds				
		10	32	100	316	1000
C/A −∞	C'/A	−4.3	−6.3	−7.7	−9.4	−10.0 db
r^2 .00	r^2	.27	.19	.15	.10	.09
z .00	z	.58	.47	.41	.32	.31
C/A −20 db	C'/A	−3.7	−7.9	−8.8	−9.2	−10.0 db
r^2 .01	r^2	.30	.14	.12	.11	.09
z .10	z	.61	.39	.36	.35	.31
C/A −10 db	C'/A	−4.0	−4.8	−6.1	−7.0	−7.0 db
r^2 .09	r^2	.28	.25	.20	.17	.17
z .31	z	.59	.55	.48	.44	.44
C/A −5 db	C'/A	−0.8	−1.3	−2.1	−2.8	−3.4 db
r^2 .24	r^2	.47	.42	.38	.34	.32
z .54	z	.84	.77	.72	.67	.64
C/A 0 db	C'/A	3.3	2.6	1.7		1.3 db
r^2 .5	r^2	.67	.64	.60		.57
z .88	z	1.15	1.10	1.03		.99
C/A +5 db	C'/A	7.8	7.0	6.6		6.0 db
r^2 .76	r^2	.86	.83	.82		.80
z 1.34	z	1.64	1.54	1.50		1.44
C/A +10 db	C'/A	14.3		12.0		11.7 db
r^2 .91	r^2	.96		.94		.92
z 1.87	z	2.30		2.09		1.94
C/A +15 db	C'/A	37.0		18.7	17.2	16.7 db
r^2 .97	r^2	.(3)8*		.983	.981	.979
z 2.43	z	5.0		2.73	2.65	2.62
C/A +20 db	C'/A	—	25.8	21.9	21.7	21.2 db
r^2 .990	r^2	—	.997	.994	.993	.992
z 2.95	z	—	3.60	3.25	3.17	3.13
C/A +25 db	C'/A	—	41.0	32.3	32.0	31.6 db
r^2 .997	r^2	—	.(4)2[b]	.(3)4[a]	.(3)3[a]	.(3)2[a]
z 3.58	z	—	5.8	4.4	4.4	4.3

[a] .(3)8 .9998.
[b] .(4)8 .99998.

are presented in Fig. 2 in terms of the Δr^2 measure. The lower abscissa is the C/A ratio in decibels; the upper abscissa is the equivalent correlation (squared). The parameter on the curves is the display duration. In terms of this measure, sensitivity changes little until a reference correlation (squared) of about 0.5 with rapid improvement, thereafter, for higher reference correlations.

Another measure of sensitivity is

the difference between the correlations in terms of a metric which normalizes the sampling variance of the correlation coefficient. If successive samples are, for example, obtained from a population correlational distribution with a correlation of near 1.0, the individual sample correlations will all be near 1.0. However, if successive samples are obtained from a population correlational distribution with a correlation of nearly zero, the sample correlations may vary over a wide range. That is, the sampling error is much wider for correlations near zero than for correlations near 1.0. The Fisher z transformation attempts to equalize the sampling variance for all correlations. In terms of the Δz measure, over most of the reference correlations, the range of scores is considerably reduced compared with the range of scores in terms of Δr^2 or Δr. We tentatively interpret this result to indicate that S's task is to sample the correlational distributions and to make a test between alternative statistical hypotheses. That performance is also reduced with shorter durations also lends support to the notion that S is acting as a tester of alternative statistical hypotheses.

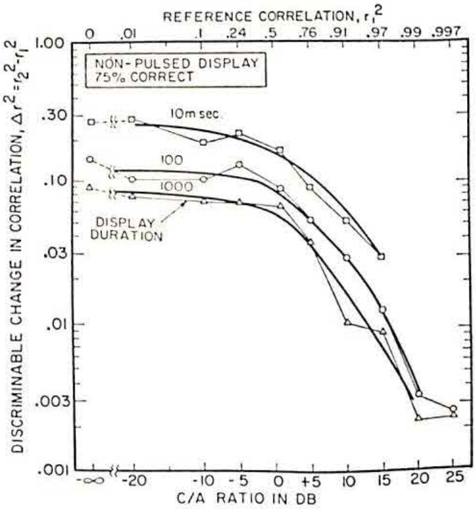


FIG. 2. Identification of two visual correlations. The upper abscissa is the reference correlation (squared); the lower abscissa is the equivalent C/A ratio in decibels. The ordinate is the difference between the required correlation (squared) for 75% correct identification and the reference correlation (squared). The ordinate is scaled logarithmically. The parameter is the display duration in milliseconds. Tests of Series II.

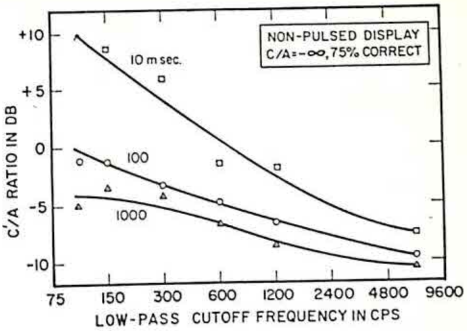


FIG. 3. Effect of low-pass cutoff frequency upon identification of two visual correlations. The abscissa is the low-pass cutoff frequency in cycles/second. The ordinate is the C'/A ratio in decibels required for 75% correct identification from a C/A ratio of $-\infty$. The parameter is the display duration in milliseconds. Tests of Series III.

Series III: Effect of Frequency Range

The effect of low-pass filtering of the noise distributions was examined for a zero reference correlation for three durations: 10, 100, and 1000 msec.[3] In terms of the stimulus to S, the effect of low-pass filtering is to thin out the density of the correlational distribution upon the oscilloscope. This operation may be alternatively considered from a sampling viewpoint as a reduction of the sample size available for inspection. In terms of the testing of alternative statistical hypotheses, we would there-

[3] The equivalence of the correlation (squared) with the ratio of the noise powers may be questioned for restricted bands of noise.

fore expect poorer discrimination with successively deeper cuts in the noise frequency distribution.

The results of the filtering tests are presented in Fig. 3. The abscissa is the low-pass cutoff frequency. The ordinate is the C'/A ratio necessary to achieve 75% correct identification. The parameter is the duration of the correlational display. The reference correlation was zero.

Larger C'/A ratios are required as the frequency range is reduced. The effect is most marked with the display of shortest duration.

Series IV: Effect of Size

If a given correlation is established, a change in size of one or both axes will leave the correlation invariant. If *S*'s task is primarily that of identifying differences in correlation, we would expect little change in identification with changes in size (until acuity considerations become important).

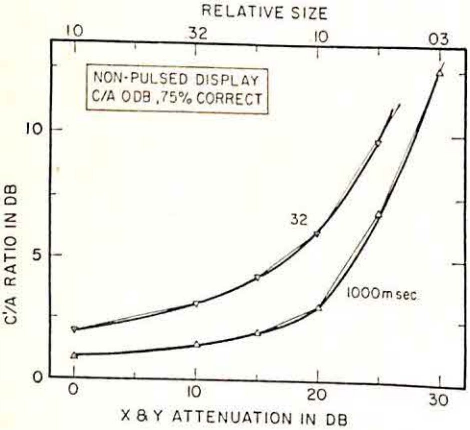


FIG. 4. Effect of size of scatterplot upon identification of two correlations. The lower abscissa is the attenuation in decibels applied to both coordinates of the oscilloscope. The upper abscissa is the relative linear size of the resulting scatterplots. The ordinate is the C'/A ratio in decibels required for 75% correct identification from a reference C/A ratio of 0 db. The parameter is the display duration in milliseconds. Tests of Series IV.

The effect of the size of the scatterplot is presented in Fig. 4. The ordinate is the C'/A ratio necessary for 75% correct identification from a C/A ratio of 0 db. The lower abscissa is the attenuation, in decibels, imposed upon both the x and y coordinates of the oscilloscope. The upper abscissa presents the relative size of the scatterplot. The parameter is the duration of the display.

Under the conditions of testing, 5- to 10-fold linear reductions in size were easily tolerated, but substantially higher C'/A ratios were required for larger size reductions.

Series V: Comparison of Positive and Negative Correlations

A further comparison between auditory and visual tests is furnished by the identification of positive vs. negative correlations. In the auditory tests (Pollack & Trittipoe, 1959), it was observed that only limited use could be made of the direction of the interaural correlation in the vicinity of zero correlation. Specifically, it was found that the absolute value of the correlation necessary to achieve a fixed identification level, $|r_i|$ was nearly identical under the following three testing procedures: Procedure A ($+r_i$ vs. r_o or zero correlation); Procedure B ($-r_i$ vs. r_o) and Procedure C ($+r_i$ vs. $-r_i$).

With visual correlations, nearly complete contribution of the direction of the correlation might be expected. That is, identification of a positive tilt of the correlational scatterplot from a negative tilt should be substantially better than identification of a positive tilt from a zero tilt. Series V tests this point.

In all, four procedures were examined in Series V. Procedures A, B, and C duplicated the corresponding test conditions for the auditory tests.

And, since Procedures A and B yielded equivalent results, only the average of their results will be presented as Procedure A:B. In addition, Procedure D($\pm r_i$ vs. r_o) was examined to determine whether uncertainty of the direction of the correlation would retard performance. In Procedure D in a random order, r_o was presented upon one-half of the trials, $+r_i$ was presented upon one-quarter of the trials, and $-r_i$ was presented upon one-quarter of the trials. Only two responses were available to the S: "stubbier" (for r_o) and "thinner" (for r_i, irrespective of the direction of the correlation).

Each procedure was examined at three durations: 10, 100, and 1000 msec. Table 2 presents the results of Procedures A:B, C, and D of Series V. The first column presents the noise burst duration, in milliseconds. The second column represents interpolated identification levels obtained from appropriate psychometric functions. The third, fourth, and sixth columns present the r_i^2 required for the stated level of identification for Procedures A:B, C, and D, respectively. (The initial decimal point has been omitted.) For example, Table 2 says that for 100-msec. noise bursts in Procedure C, $+r^2$ of .045 when paired with $-r^2$ of .045 yielded 75% correct identifications. The ratio between the corresponding entries for Procedures A:B and C is presented in the fifth column.

TABLE 2

SUMMARY OF RESULTS WITH POSITIVE AND NEGATIVE CORRELATIONS

Duration Msec.	% Corr.	Procedure			
		A:B	C	$\frac{A:B}{C}$	D
10	95	376	237	1.6	410
	90	295	184	1.6	343
	85	245	143	1.7	285
	80	205	112	1.8	237
	75	174	089	2.0	195
	70	143	072	2.0	161
	65	119	056	2.1	122
	60	097	025	3.9	089
100	95	195	137	1.4	192
	90	155	087	1.8	152
	85	127	065	2.0	130
	80	104	054	1.9	112
	75	084	045	1.9	097
	70	067	035	1.9	081
	65	046	026	1.8	063
	60	029	017	1.7	043
1000	95	164	074	2.2	149
	90	124	049	2.5	125
	85	099	038	2.6	097
	80	075	031	2.4	082
	75	056	026	2.2	069
	70	040	022	1.8	056
	65	024	018	1.3	045
	60	013	013	1.0	033

Table 2 makes three points: (*a*) Procedure C (positive vs. negative correlation) yields the highest sensitivity in that the required correlational differences are substantially lower than obtained with the other procedures. (*b*) The ratio of the correlations (squared) yielded by Procedures A:B and C in the fifth column of Table 2, is nearly 2 to 1, especially for intermediate levels of identification. This result suggests the correlation (squared) measure is an appropriate measure of the correlational variable in that it is additive about zero correlation. (*c*) Uncertainty of direction of correlation associated with Procedure D has a small, but significant, effect in reducing accuracy of identification.

Series VI: Effect of Sampling Pulse Frequency

Series I–V considered a correlational display in which a continuously moving electron stream defined the distribution. Series VI–VIII considered the equivalent noise distributions but with the trace visible only in terms of discrete dots spaced in

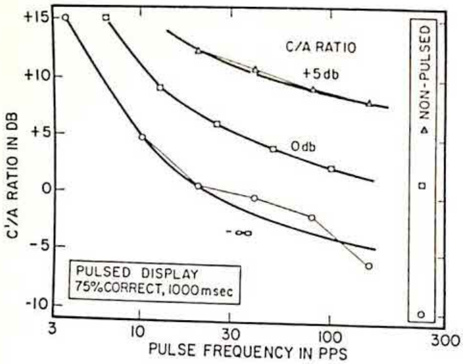


FIG. 5. Effect of discrete pulsing frequency upon identification of two correlations. The abscissa is the pulsing frequency, in pulses/second. The ordinate is the C'/A ratio in decibels required for 75% correct identification from each of three reference C/A ratios (listed as the parameter). For comparison, the results of Table 1 for the nonpulsed display of 1000 msec. are presented. The duration of the display was 1000 msec. Tests of Series VI.

time. This was accomplished by briefly pulsing the z axis of the oscilloscope. The number of samples from the distribution is equal to the number of pulses delivered to the z axis within the available display interval. The number of samples is the product of the pulsing rate (impulses per second) and the available display duration (in seconds). In terms of a statistical hypothesis tester, S's performance would be expected to improve as the size of the sample is increased.

Figure 5 presents the effect of varying the pulsing rate with a display of available duration constant at 1.0 sec. For comparison, results under corresponding nonpulsed conditions are also presented from Fig. 2.

Performance improves substantially with higher pulsing frequencies.

Series VII: Effect of Sampling Pulse Number

The total number of samples, as represented by the total number of pulses per observation interval, is determined by the product of the pulsing frequency and the available duration of the display. Which of the three parameters—number of pulses, pulsing frequency, or available duration of the display—is the critical determinant of performance?

Figure 6 presents the results of tests carried out over a wide range of available display durations and pulsing rates. The abscissa is the available duration of the display; the ordinate is the C'/A ratio in decibels required for 75% correct identification; the parameter is the number of samples (pulses) within the display duration. For purposes of comparison, the corresponding results for the nonpulsed display from Fig. 2 are also presented.

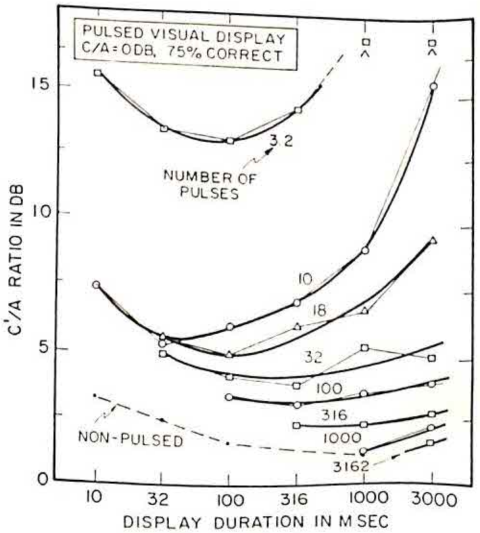


FIG. 6. Effect of number of discrete samples upon identification of two correlations. The abscissa is the display duration in milliseconds. The ordinate is the C'/A ratio required for 75% correct identification from a C/A ratio of 0 db. The parameter is the number of samples produced by pulsing the z axis of the oscilloscope. For comparison, the results of Table 1 for the nonpulsed display are presented as the dashed curve. Tests of Series VII.

Over a wide range of conditions, performance is nearly invariant with a constant number of samples per presentation.[4] The notable exception is the rapid deterioration of discrimination for displays with a small number of pulses. With this exception in mind, we conclude that identification of pulsed correlational displays is, in large part, independent of the display duration and pulsing rate, except insofar as these factors determine the total number of sampling pulses.

Series VIII: Effect of Sampling Pulse Duration

The number of sampling pulses was considered to be the critical determinant of performance in Series VII. However, it may be noted that the effect of sampling pulse number is confounded by the sampling pulse duration. (The total sampling duration is the product of the total number of sampling pulses by the duration per sampling pulse.) In order to test for confounding, additional identification tests were carried out with sampling pulse durations from 2 to 80 μsec with sampling pulse numbers of 3 to 80 pulses per display. It may be noted that, under the longer sampling pulse durations, large segments of the correlated 100–6800 cps wave forms were obtained rather than the discrete dots observed with the shorter pulsing durations.

Over the range of pulse durations examined, equivalent performance was observed for a constant number of sampling pulses, irrespective of the sampling pulse duration. This result lends support to Series VII which concluded that the number of sampling pulses was the critical determinant of performance.

DISCUSSION

The deterioration in the identification of correlational scatterplots with reduction in (*a*) display duration, (*b*) noise bandwidth, and (*c*) pulsing number, all lend support to the interpretation of *S*'s task as that of testing alternative statistical hypotheses.

In this light, the deterioration of performance is not to be attributed to a change in *S*'s performance, but rather to a reduction in the reliability of the statistical information available to the display. Under comparable conditions, the performance of an ideal statistical observer would also suffer. Supporting evidence that statistical sampling of the stimulus was often a controlling factor upon performance is furnished by the extremely high inter-*S* agreement with short nonpulsed samples or with a small number of pulses. Under these circumstances, all *S*s were often in perfect agreement in selecting between the two correlations and often incorrectly.

We have examined several models based on the sampling distribution of the correlation coefficient. To date, we have been singularly unsuccessful in obtaining a model which yields satisfactory quantitative predictions of the data. We strongly suspect that the theoretical development of a performance model for correlational distributions for the ideal statistical observer will point the way to the required model of the human observer.

One final point: it is an interesting exercise to inquire about an appropriate representation of the correlational variable within a discrimination space. Distance in discrimination space is scaled in terms of the confusability among the items: the greater the confusability between two conditions, the closer is their representation in discrimination space. This approach suggests that we might profitably search for an appropriate metric of confusability such as the d' measure of the theory of signal detectability.

---

[4] The rapid stabilization of the correlation between 4 and 36 sampling pulses is illustrated by Licklider and Dzendolet (1948, Fig. 2C, p. 122).

An alternative approach may be employed for two special cases. If the absolute value of correlations required for a given level of identification $|r_i|$ is identical in the testing of ($+r_i$ vs. r_o or zero correlation); ($-r_i$ vs. r_o); and ($+r_i$ vs. $-r_i$), we have three points equidistant in discrimination space: $+r_i$, $-r_i$, and r_o. The appropriate representation is two discrimination axes, labeled $+r_i$ and $-r_i$ terminating at r_i and separated by an angle of 60°. This result is, in fact, obtained with interaural auditory noise correlations. Direction of correlation may be discriminated, but its effects are only partially additive.

Another special case is observed when the effect of the direction of correlation is completely additive. That is, with respect to an appropriate metric, we need twice the value of the metric required for ($+r_i$ vs. r_o) or ($-r_i$ vs. r_o) as for ($+r_i$ vs. $-r_i$). In the present examination of the visual correlational scatterplot, r_i^2 appears to be the appropriate metric, at least for intermediate identification levels. The representation of the visual scatterplot in discrimination space is, then, simply a single discrimination continuum, scaled in terms of r^2, and symmetrical about r_o.

SUMMARY

Visual correlation scattergrams were obtained by mixing a common noise source, C, with independent noise sources, A and B, and displaying the mixtures $(C + A)$ and $(C + B)$ across the x and y coordinates of an oscilloscope. The correlational distributions were displayed upon the oscilloscope either (a) by a continuous beam of electrons or (b) by sampling the beam to obtain discrete samplings of the display. The task of S was to identify whether the reference correlation or a higher correlation was presented. The variables examined were the duration, size, and sampling characteristics of the display and the upper cutoff frequency of the noise distributions.

In terms of the magnitude of the difference between the reference and variable correlations required for 75% correct identification: (a) Sensitivity is greatest for the highest reference correlations. (b) Sensitivity decreases with narrower bandwidths and shorter display times, with rapid changes below burst durations of 100 msec. (c) Sensitivity for the sampled discrete pulsed display is determined by the number of sampling pulses, and is independent of the pulsing rate and duration of the display, except insofar as they determine the number of sampling pulses.

The results are interpreted in terms of the task of S as a tester of alternative statistical hypotheses under conditions of varying reliability of the display information.

REFERENCES

LICKLIDER, J. C. R. The influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Amer.*, 1948, **20**, 150–159.

LICKLIDER, J. C. R., & DZENDOLET, E. Oscilloscopic scatterplots illustrating various degrees of correlation. *Science*, 1948, **107**, 121–124.

POLLACK, I., & TRITTIPOE, W. Binaural listening and interaural noise correlation, *J. Acoust. Soc. Amer.*, 1959, **31**, 1250–1252. (Abstract)

ZERLIN, S. The interaural time disparity threshold as a function of the interaural correlation. *J. Acoust. Soc. Amer.*, 1959, **31**, 127. (Abstract)

(Received June 15, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

TASK DIFFICULTY AND DRIVE IN VERBAL LEARNING

R. R. STANDISH AND R. A. CHAMPION

University of Sydney

In view of the wide disparity in the results of experiments on the effect of variation in drive strength in verbal learning (e.g., Jensen, 1958) the following study was made in an attempt to test the basic essentials of Spence's theory (Spence, 1958) under the least equivocal conditions. A situation was first provided in which the correct response was clearly dominant, the prediction being that the performance of high-drive *S*s would be superior to that of low-drive *S*s with medium-drive *S*s occupying an intermediate position. It was thought that some previous *E*s in their selection of material to be learned may have merely demanded responses which were relatively easy to acquire without ensuring the actual dominance required by the theory. For example, if a "high association" or easy response is given on 20% of a block of trials then it must have been dominated at least momentarily by some incorrect response on 80% of the block of trials. Thus an increase in drive strength might impair the performance of this "easy" response as much as that of a "difficult" response given on only 10% of the same block of trials. The present interpretation therefore favors clear dominance of the correct response rather than relative ease or high association value if the performance of a high-drive group is to be superior.

A second type of situation was devised to ensure that an incorrect response was clearly dominant at the outset of acquisition, the prediction in this case being that the performance of low-drive *S*s would be superior to that of high-drive *S*s up to the point where the clear dominance of the incorrect response was lost, when there would be a reversion to the first type of situation. As before, medium-drive *S*s were expected to occupy an intermediate position. Judging by the results of some other experiments (e.g., Spence, 1956) in which low-drive *S*s have performed better on low association-value or "difficult" responses, this second type of situation is more readily obtainable. The above two predictions are most simply taken together as involving an interaction between task difficulty and drive level.

METHOD

Subjects.—The *S*s were 63 students taking courses in psychology or medicine at the University of Sydney. Three *S*s obtained a "lie score" of 7 or more on the index of drive level, the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953), and were rejected. The remaining 60 *S*s were each allocated to one of three drive groups on the basis of their anxiety score on the MAS. The 20 low scorers (2–11) constituted the low-drive group (Mean = 7.5, Mdn. = 7.5, $SD = 2.4$), the 20 high scorers (20–33) the high-drive group (Mean = 24.8, Mdn. = 24.0, $SD = 3.4$) and the remaining 20 *S*s made up the medium-drive group (Mean = 15.5, Mdn. = 15.5, $SD = 2.2$).

Learning materials.—A list of 10 easy paired-associates was constructed by using common associates in the Kent-Rosanoff list (Rosanoff, 1938). To reduce intralist confusions no stimulus word was used which had been given as a response at any point in the Kent-Rosanoff study; the response word in each case was the answer most commonly given in free association. To ensure clear dominance of the correct response, performance on this easy list was not investigated until each *S* had given the correct response on two successive trials and no measures of performance (response speed) were used prior to this point. A list of 10 difficult paired-associates was constructed by taking the stimulus words of the easy list and pairing them with response words not given in free association according to the Kent-Rosanoff list. To ensure clear dominance of the incorrect response all *Ss* learned both lists, with the difficult list following the easy list.

Procedure.—The items were presented in a Ranschburg exposure apparatus with approximately 30 sec. between trials. The *S* was allowed 4 sec. to anticipate the response with the stimulus word present (in blue), and the response word (in red) was then presented for 4 sec. To reduce intralist associations the items of each list were presented in three orders, repeated as necessary. An electric commutator actuated the exposure apparatus and started an electronic chronoscope simultaneously with the presentation of each stimulus word. The chronoscope was stopped by *E* pressing a key as soon as *S* gave the correct response or else it was reset after 4 sec. in readiness for the next stimulus. The easy list was presented to each *S* first, to the criterion of two successive errorless trials followed by eight further trials. Then the difficult list was presented until the same criterion had been reached. Finally, each *S* answered the MAS items.

RESULTS

The results obtained with speed of correct response (Reciprocal Latency $\times 10^5$) as a measure of performance are presented in Fig. 1. The scores plotted for performance on the easy list are taken from the first two errorless trials and the following eight trials, all of which are labeled as Trials 1–10. The scores for the difficult list are taken from two stages of learning: (*a*) the first 10 trials (labeled also as Trials 1–10) and (*b*) the first two errorless trials and the following eight trials (labeled as Trials A–J). It should be noted that the mean scores for the first two trials on the difficult list are spuriously depressed through the use of an arbitrary maximum response time of 4 sec.

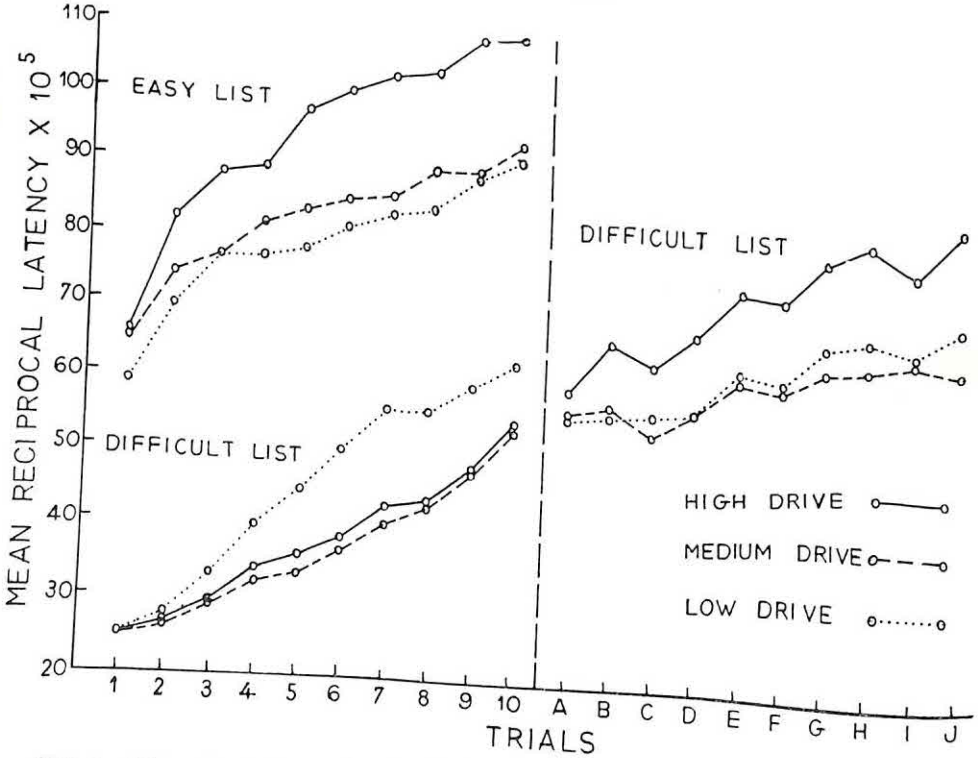


FIG. 1. Mean response-speed early (Trials 1–10) and late (Trials A–J) in learning.

Trials 1–10 as defined above provide the primary data of the experiment. These data are in accord with the prediction for the easy list, with the high-drive group superior in performance to the low-drive group throughout and the medium-drive group falling between these extremes. On the difficult list the prediction is confirmed to the extent that the low-drive group shows superior performance to the high-drive group on all trials except the first, but there is a deviation from prediction in that the high-drive *S*s perform at a consistently higher level than the medium-drive *S*s when the opposite result was expected. The results of an analysis of variance in these data are summarized in Table 1, homogeneity of variance having been established. The predicted interaction is statistically significant.

A further test of the general theory under consideration is available in the data for the difficult list over Trials A-J as defined above and in a comparison of these data with the scores from Trials 1–10 for the same list. When correct responses are given on the difficult list it takes on the character of the easy list, leading to the prediction that high-drive *S*s should now perform at a superior level despite their previous inferiority. That this effect does occur is shown in Fig. 1 although now the low-drive group is slightly superior to the medium-drive group at most stages whereas it would have been expected to be inferior in performance. The effect now under consideration in the comparison of performance on the difficult list at two stages of learning involves an interaction similar to that already considered in Table 1. This was tested in a second analysis of variance, summarized in Table 2. The interaction is statistically significant and the second analysis confirms the finding in the primary data.

TABLE 1

Analysis of Variance in Speed of Response Scores over Trials 1–10

Source	*df*	*MS*	*F*
Task difficulty (A)	1	588,699.18	488.38**
Drive level (B)	2	4,148.01	3.44*
A × B	2	15,655.73	12.99**
*S*s	114	1,205.42	6.90**
Measures	960	175.69	

* $P < .05$.
** $P < .001$.

TABLE 2

Analysis of Variance in Speed of Response Scores on the Difficult List in Early and Late Stages of Learning

Source	*df*	*MS*	*F*
Stage of learning (A)	1	183,300.84	228.85**
Drive level (B)	2	4,838.81	6.04*
A × B	2	10,117.41	12.63**
*S*s	114	800.95	5.54**
Measures	960	144.51	

* $P < .01$.
** $P < .001$.

A final test of the theory was made with two other measures of performance available in the acquisition of the difficult list, percentage fre-

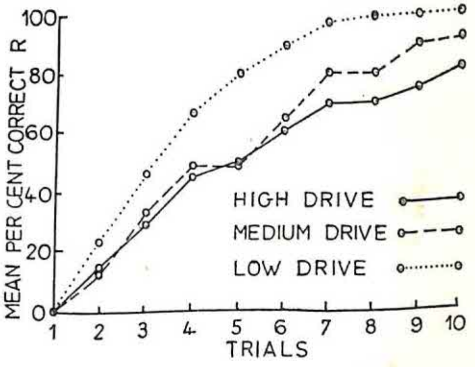


Fig. 2. Mean percentage frequency of correct response for the difficult list early in learning.

quency of correct response and number of trials to the criterion. The result using the first of these measures over Trials 1–10 is plotted in Fig. 2, wherein it may be observed that the relative disposition of the low-, medium-, and high-drive groups deviates only slightly from that predicted. The mean trials-to-criterion scores for the low-, medium-, and high-drive groups were 7.86, 12.25 and 14.85 trials, respectively, and exactly in accord with the prediction.

Discussion

The results of this experiment are clearly compatible with the theory under consideration, presumably because it was conducted as nearly as possible within the specified boundary conditions. It would therefore seem that the effect of other methods of varying task difficulty in verbal learning, e.g., through intralist similarity, should not necessarily be accepted as support for or disproof of the theory in the present state of knowledge but that an attempt should rather be made to extend the theory to account for more complex circumstances. Some features of the present study suggest factors which might be taken into account in this endeavor.

First, the equivocal results of the medium-drive group deserve attention. These *S*s fell with the high-drive group when the incorrect response was dominant (difficult list, Trials 1–10) and with the low-drive group when the correct response was dominant (easy list, Trials 1–10 and difficult list, Trials A-J) whereas it had been expected that they would occupy an intermediate position throughout. The reason for this finding may lie in the nature of the MAS as a measuring instrument in need of further refinement.

Secondly, there appear to be some discrepancies in the data according to the measure of performance used. With speed of correct response as an index of response strength the performance curves for the acquisition of the difficult list cross at an intermediate value of about 60 units on the scale used (by interpolation in Fig. 1). With percentage frequency of correct response as a measure, however, no such crossing was observed (Fig. 2). An additional phenomenon is the drop in the response speed of the low-drive group on the difficult list at about the stage of the first four errorless trials (Trials A-D, Fig. 1) whereas the other two groups continue to improve; this finding was not apparent with the other measure of performance. These discrepancies are almost certainly due to the fact that response speed is a more sensitive measure and one which does not present an arbitrary maximum level, but it is possible that percentage frequency of correct response involves some complication as yet not made explicit.

Finally, whereas performance on the easy task at Trial 10 when averaged over the three drive groups is 98.3 units on the scale of response speed, that for the difficult task at Trial J is only 73.7 units on the same scale (Fig. 1). The similar disposition of the high- and low-drive groups at these points, as well as the use of the same criterion of performance, would suggest that the absolute level of performance should be the same in the two cases and the obtained difference was not anticipated. A reason for expecting a difference may be found after the event in the possible inhibiting effects of nonreinforcement (Spence, 1956) which might carry over from Trials 1–10 to Trials A-J in the case of the difficult list but which could not be generated in Trials 1–10 with the easy list.

Summary

Situations were sought in which the basic implications of Spence's theory about the interaction of task difficulty and drive in verbal learning could be tested in the least equivocal fashion. An easy list in which the correct response clearly dominated was obtained by measuring speed of response over a series of errorless trals and a difficult list involving clear dominance of the incorrect

response was achieved by having *S*s then learn new responses to the stimuli of the easy list. With score on the Taylor MAS as an index of drive level, high- and low- drive groups performed as expected, with the high-drive group superior on the easy list and inferior on the difficult list but medium-drive *S*s failed to occupy the predicted intermediate position with any consistency. The nature of the MAS as a measuring instrument, the index of response-strength used, and inhibition due to nonreinforcement were indicated as three specific factors contributing to disparity in the results of experiments in this area as well as general failure to keep within the boundary conditions of the type of theory being tested.

REFERENCES

JENSEN, A. R. Personality. *Ann. Rev Psychol.*, 1958, **9**, 295–322.

ROSANOFF, A. J. *Manual of psychiatry.* New York: Wiley, 1938.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received June 19, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

ASSOCIATION VALUE AND FAMILIARITY IN SERIAL VERBAL LEARNING[1]

RICHARD H. LINDLEY

Carnegie Institute of Technology

A well known generalization about verbal learning is that serial lists of high association value (HAV) nonsense syllables are easier to learn than lists of low association value (LAV). This generalization holds true under fairly varied conditions; e.g., when the methods of complete presentation and unaided recall are used (McGeoch, 1930) or when the syllables are spelled and presented serially on a memory drum (e.g., Underwood & Richardson, 1956).

Analysis of lists of nonsense syllables in which association values were determined (Glaze, 1928; Hull, 1933; Krueger, 1934) shows that some of the syllables are familiar since they are the first three letters of English words (e.g., BEL-BELL; FAS-FAST). Ninety-seven per cent of the 100% association value syllables in Glaze's lists are the first three letters of one or more of the 20,000 most common words in the English language as determined by Thorndike and Lorge (1944). The comparable figures for the 80% and 0% association value syllables are 70% and 0%, respectively. Thus differences in the rate of learning HAV and LAV syllables may simply be due to differences in the familiarity of these kinds of items.

The present study compared the rates of learning of lists of nonsense syllables which had the same published association values but which differed appreciably in familiarity. If the relationship between association value and rate of learning is due to familiarity, the familiar syllables should be easier to learn than the unfamiliar ones regardless of association value.

A second implication of the familiarity hypothesis is that nonsense syllables should be no more difficult to learn than meaningful words provided that both kinds of items are equally familiar. McGeoch (1930) found that words were easier to learn than HAV syllables; however, no control over familiarity was reported in that study. The present study controlled for familiarity.

METHOD

Experimental design.—There were four main groups of *S*s. Group I learned a list of meaningful words; Group II learned familiar HAV syllables; Group III learned unfamiliar HAV syllables; and Group IV learned unfamiliar LAV syllables. The study consisted of two replications; the first used syllables drawn from Glaze's lists and the second used syllables drawn from Krueger's lists. Twelve *S*s were assigned to each group in each replication, making a total of 24 *S*s in each main group.

In the first replication the meaningful words were constructed by adding letters to the familiar HAV syllables to form words (e.g., FOL-FOLD). The results showed almost no difference in the rate of learning of the familiar HAV syllables and words. The failure to obtain a significant difference could have been due to the fact that the words were longer than the syllables; therefore, *S*s in Group I in the first replication were replaced with *S*s who learned words of the same length

---

[1] This paper was presented at the Midwestern Psychological Association Convention in 1959. The research was supported in part by a grant from the Humanistic and Social Studies Division of Carnegie Institute of Technology. Thanks are due Mary Ann Turnquist for clerical assistance.

TABLE 1

LISTS USED

Words (Repl. 1 and 2)	Glaze's Syllables (Repl. 1)			Krueger's Syllables (Repl. 2)		
Group 1[a]	Group 2[a]	Group 3	Group 4	Group 2[a]	Group 3	Group 4
LIP	WID	RUJ	LEQ	WES	LUV	YUT
TAX	FOL	WIM	CIJ	BUR	QIT	BIW
DOG	JAC	BOC	WUB	PIC	ROX	GOX
RED	KEP	DEY	ZOM	GOV	WUN	DAJ
CUT	BIR	LAQ	DAX	NAM	DAC	ZIQ
MAN	HOM	VIF	NEH	JUL	YEZ	FAP
FEW	SUG	TAZ	VAF	DIF	KOS	KEC
JOY	NEX	NOK	YOP	YEL	JEM	VUH
HIS	GAV	HUP	GUK	TAK	FAX	MOJ

[a] All items for Groups 1 and 2 occurred 100 times per million or more in the Thorndike-Lorge count except TAX, JAC, and JUL which occurred 50 times per million but less than 100 times.

as the nonsense syllables. The latter *S*s were run during the second replication.

Materials.—Familiarity was estimated by using the G or General count of the Thorndike and Lorge (1944) word count. The familiar syllables were the first three letters of English words which occurred 50 times per million or more in English text. The unfamiliar syllables were not the first three letters of any words occurring one or more times per million. Thirteen of these syllables were not parts of any words which occurred 10 times per million or more; five of the syllables occurred once or twice in positions other than the first three letters in words which occurred 10 times per million or more. Thus the familiar HAV syllables were clearly more familiar than the unfamiliar HAV syllables. The meaningful words were selected from three-letter words in the Thorndike-Lorge count which occurred 50 times per million or more.

Each list contained nine items. The items were selected with the restriction that formal intralist similarity be low. No consonant letter appeared more than twice in any one list. The items are presented in Table 1 in the orders in which they were learned by *S*s. The mean association values and ranges of Glaze's syllables for Group II were 84.6% and 67%–100% respectively; for Group III, 84.6% and 67%–100%; for Group IV, 7.4% and 0%–27%. The corresponding values of Krueger's syllables for Group II were 94.5% and 85.5%–99%; for Group III, 94.4% and 89.5%–96%; for Group IV, 49.4% and 34%–60%.

Procedure.—The items were presented on a memory drum at a 2-sec. rate. An asterisk served as a cue for the first item. There was a 6-sec. pause between trials. Typical instructions for learning by the anticipation method when the items are spelled were read to *S* at the start of the experiment. The lists were learned for 25 trials and the mean number of correct responses occurring during these trials constitute the major data of this study.

At the termination of serial learning, *S* was instructed to report any meaningful associations he had made to the nonsense syllables during serial learning. The *E* spelled the syllables one at a time and *S* reported any associations he had made.

Subjects.—The *S*s were 96 undergraduate men and women from Carnegie Institute of Technology. None of the *S*s had had previous nonsense syllable learning experience. The *S*s were assigned to groups according to a randomized block design.

RESULTS

Serial learning.—The means and *SD*s for the mean number of correct responses during the 25 serial learning trials are presented in Table 2. Analysis of variance showed that only the differences between groups were significant ($F = 19.90$, $df = 3$ and 66, $P < .01$). Since the F for replications was not significant, the means for the individual groups were pooled and the differences between the main groups were tested

TABLE 2

MEANS AND *SD*s FOR NUMBER OF CORRECT RESPONSES IN 25 TRIALS

Groups	Replication 1		Replication 2	
	Mean	*SD*	Mean	*SD*
I	166.25	15.69	169.25	22.89
II	153.67	45.43	151.83	34.48
III	116.17	37.80	123.25	51.80
IV	103.67	30.95	83.00	27.18

TABLE 3

MEAN NUMBER CORRECT PER TRIAL FOR SUCCESSIVE BLOCKS OF FIVE TRIALS FOR THE MAIN GROUPS

Groups	Trial Blocks				
	1–5	6–10	11–15	16–20	21–25
I	3.49	5.93	7.33	8.30	8.34
II	3.26	5.45	6.61	7.28	7.77
III	2.32	3.99	4.80	6.18	6.69
IV	1.44	2.63	3.80	4.98	5.85

with the *t* test; a difference of 22.01 is significant at the 5% level. Thus all differences reached $P = .05$ except that between Groups I and II.

Inspection of Table 2 shows that there is a trend for the meaningful words to be easier to learn than the familiar HAV syllables. Further analysis of the data showed that four *S*s in Group II failed to learn the list to a criterion of one perfect recitation within 25 trials whereas all *S*s in Group I learned the list within 25 trials. When medians are compared, the difference between Groups I and II disappears. (The median number correct was 166 for Group I, 163 for Group II, 124.5 for Group III, and 93.5 for Group IV.) In terms of the number of trials to reach a criterion of one perfect recitation, the mean number of trials was 11.83 for Group I ($N = 24$) and 11.65 for Group II ($N = 20$). Thus, for the majority of *S*s, the familiar HAV syllables were no more difficult to learn than the meaningful words.

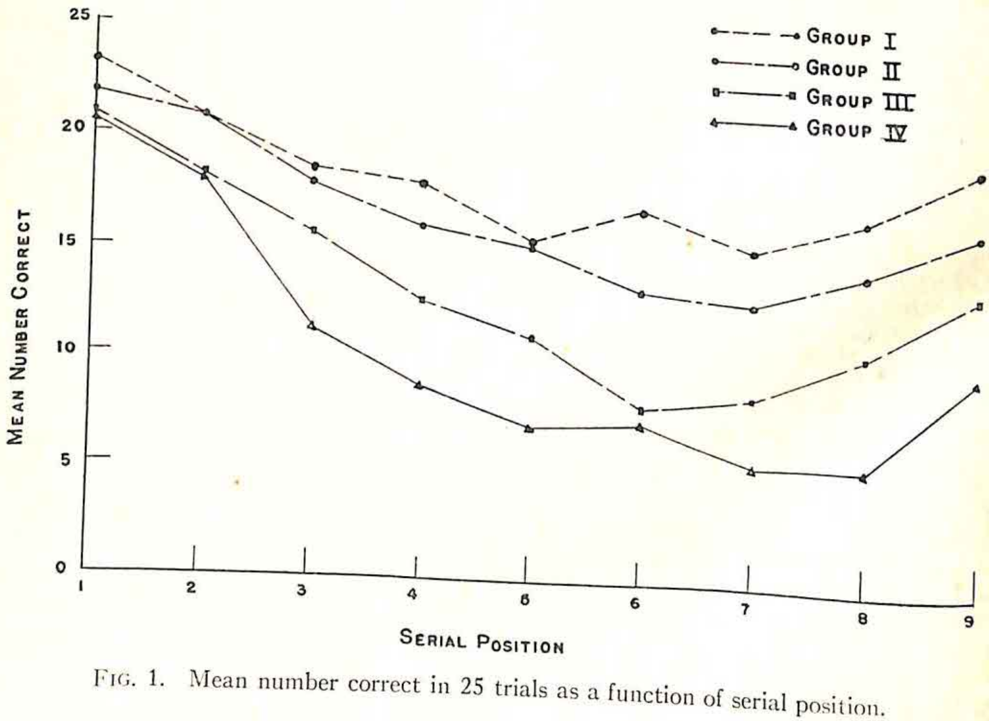


FIG. 1. Mean number correct in 25 trials as a function of serial position.

TABLE 4

MEANS AND *SD*s FOR NUMBER OF ASSOCIATIONS

Groups	Replication 1		Replication 2	
	Mean	*SD*	Mean	*SD*
II	6.33	2.49	7.00	1.78
III	4.33	2.17	5.67	1.65
IV	4.08	2.29	4.00	2.16

The mean number of correct responses per trial for successive blocks of five trials for the main groups are presented in Table 3. The curves for the high familiarity items are negatively accelerated and asymptotic to the x axis while the curves for the low familiarity items approximate straight lines. This may simply reflect the fact that the high familiarity items were better learned during the 25 trials than the low familiarity items. Figure 1 presents serial position curves for the four main groups. Skewed, bowed serial position curves which are typical of learning disconnected verbal material were obtained for all groups. The curves for the high familiarity items are less bowed than the curves for the low familiarity items.

Number of associations.—The means and *SD*s for the number of associations reported by *S*s are presented in Table 4. Analysis of variance showed that only the differences between groups were significant ($F = 9.99$, $df = 2$ and 44, $P = .01$). The t test shows that a difference of 1.23 associations between groups will be significant at $P = .05$. Thus only the differences between Groups II and III, and II and IV reach an acceptable level of significance.

DISCUSSION

The results clearly show that familiarity is an important variable in serial nonsense syllable learning. Even though the published association values of the syllables learned by Groups II and III were equal, the familiar syllables were significantly easier to learn than the unfamiliar syllables. Furthermore, when familiarity was controlled, the familiar syllables were no more difficult to learn than the meaningful words for most *S*s.

The results also showed that there was a direct relationship between ease of serial learning and the number of associations evoked by the syllables, although only the differences in associations between the familiar and unfamiliar syllables reached the .05 level of significance. These results are consistent with the theories of Mandler (1954), and Dollard and Miller (1950) which predict that items which elicit distinctive implicit responses will be easier to learn than items which do not elicit such distinctive responses. The *S*s in the present study reported more associations (distinctive responses) to the familiar syllables than they did to the unfamiliar syllables. The fact that the unfamiliar HAV syllables were easier to learn than the LAV syllables can be explained by this reasoning. That is, the unfamiliar HAV syllables elicited more associations than the LAV syllables and thus may have been more distinctive.

The results of this study are consistent with results obtained by Hall (1954). In that study familiar words were found to be easier to recall than unfamiliar ones. Thus familiarity is an important variable in the learning of both nonsense syllables and words.

SUMMARY

The present study tested the hypothesis that the relationship obtained between published association values and ease of serial learning was due to familiarity. The results showed that lists of familiar syllables were easier to learn than lists of unfamiliar syllables when the items were spelled, even though the average published association values for the lists were identical. Furthermore, no significant difference was obtained between the rate of learning of meaningful words and nonsense syllables when familiarity was equated. The average number of associa-

tions evoked by the syllables during learning was directly related to ease of learning. The results were interpreted as being consistent with theories which emphasize the role of distinctive implicit responses in verbal learning.

REFERENCES

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

GLAZE, J. A. The association value of nonsense syllables. *J. Genet. Psychol.*, 1928, **35**, 255–269.

HALL, J. F. Learning as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 138–140.

HULL, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

MCGEOCH, J. A. The influence of associative value upon the difficulty of nonsense-syllable lists. *J. Genet. Psychol.*, 1930, **37**, 421–426.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bureau Publ., Columbia Univer., 1944.

UNDERWOOD, B. J., & RICHARDSON, J. A. The influence of meaningfulness, intralist similarity, and serial position on retention. *J. exp Psychol.*, 1956, **52**, 119–126.

(Received June 23, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

EFFECTS OF PROBABILITY OF MISINFORMATION AND NUMBER OF IRRELEVANT DIMENSIONS UPON CONCEPT IDENTIFICATION[1]

VLADIMIR PISHKIN[2]

University of Utah

The primary aim of the present experiment was to investigate the effect of incorrect, or misinformative, feedback (MF), in conjunction with various degrees of task complexity, upon the performance of human *S*s in concept identification.

Variation of misinformative feedback, MF, in this experiment, was an operation similar to partial reinforcement in a two-choice guessing problem, where asymptotic performance typically reaches probability matching. The present experiment differed from guessing game studies (e.g., Grant, Hake, & Hornseth, 1951; Humphreys, 1939) in the following respects: (*a*) percentage of correct feedback for either of the two responses depended upon only one binary dimension of the stimulus, and (*b*) stimulus information irrelevant to either response was introduced in a quantifiable manner. It was hypothesized that the introduction of irrelevant information would reduce rate of learning. Despite differences in methodology, it was expected that probability matching would take place in this experiment as in other partial reinforcement studies. Morin (1955) has shown that the acquisition of correct responses in a two-choice discrimination situation is negatively and progressively influenced by MF. Morin's data suggest that various degrees of MF in a concept identification situation would produce differential acquisition rates of correct responses.

Estes and Straughan's (1954) model demonstrated that with the exception of a few deviant cases, terminal response probability approached theoretical asymptotic response probability, π, the probability of reinforcement for that response. Restle (1955, 1957) has proposed a model similar to Estes and Straughan's except that the proportion of relevant stimulus cues in a problem determines the rate of learning. This theory assumes a population of available stimuli consisting of sets of cues which are either relevant or irrelevant to the problem's solution. Formally, $\delta = r/r + i$, where r is number of relevant cues, i is number of irrelevant cues, and δ is the learning rate. By assuming that over a series of trials all irrelevant cues will be adapted and thus rendered ineffective, and that probability of response equals proportion of conditioned, unadapted cues, Restle was somewhat successful in predicting the outcome of a variety of two-choice discrimination problems. In Restle's (1955) framework, the number of errors (E_t) made before solution is approximated by:

---

[1] This article is based on a dissertation submitted to the Department of Psychology of the Graduate School of the University of Utah in partial fulfillment of the requirements for the PhD degree. The author is especially indebted to Lyle E. Bourne, Jr., for his guidance and assistance throughout the course of this investigation. This research was supported, in part, by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service.

[2] Now at Research Laboratory, Veterans Administration Hospital, Tomah, Wisconsin.

$$E_t \cong \tfrac{1}{2}+\tfrac{1}{2}\left[\frac{\log\theta}{(1-\theta)\log(1-\theta)}\right] \quad [1]$$

where θ is a rate of learning parameter, originally assumed equal to δ. In a later paper, Restle (1957) reasoned that rate of learning parameters depend on π, the proportion of time a relevant cue is reinforced. Further, Restle proposed that a relevant cue which is reinforced less than 100% of the time has validity less than unity. The following equation represents cue validity, V, as a function of π:

$$V = 4\pi^2 - 4\pi + 1 \quad [2]$$

Then, in the extended model:

$$\theta = V\delta \quad [3]$$

Since S may not perceive all of the cues presented, the parameter δ must be estimated from data. Once this has been done, E_t should be predictable for various conditions in the present design. Thus, this study serves, in part, as a test of Restle's model.

The threefold purpose of the present experiment was: (*a*) to determine the effect of MF under various degrees of complexity, (*b*) to test the hypothesis of probability matching, and (*c*) to account for the data in terms of Restle's discrimination model.

METHOD

Subjects.—The *S*s were 180 volunteer students in elementary psychology courses who were divided randomly into 45 treatment groups of 4 *S*s each (2 men and 2 women).

Task and apparatus.—All *S*s were presented a series of 192 geometric patterns with a Dunning animatic projector. The screen, mounted at *S*'s eye level, was bordered by an opaque shield which prevented *S* from seeing other than pertinent apparatus. In front of *S* was a panel with two response keys and two amber signal lamps. The *S*'s task was to solve a two-choice discrimination problem by successively matching one or the other of two keys to invariant pattern dimensions, e.g., depression of the left key for the image of a triangle; of the right, for a square, etc. The keys were not marked in any way.

Procedure.—At the beginning of each session, *S* was verbally instructed as to the nature of the task, the meaning of the feedback lamps, number of trials required, and manipulation of the controls. The *S* began his task by viewing a geometric pattern projected on the screen. After a self-determined interval, *S* responded by pressing one of the two keys. The momentarily depressed key initiated a signal which recorded *S*'s choice of responses on the chart of an Esterline-Angus operations recorder and triggered a cascaded multiple-phase electronic timer.

Phase 1 of the timer advanced the filmstrip to a blank frame. Phase 2 activated a Western Union tape transmitter precoded with a tape which matched the filmstrip programming, thereby enabling the information feedback, irrespective of *S*'s choice, to be registered on the Esterline-Angus operations recorder. In addition, the same information coded on the tape led to the illumination of one of the two panel lamps above *S*'s keys for a duration of 1 sec. Phase 3 advanced the filmstrip to the next stimulus permitting *S* to begin another trial 5 sec. after *S*'s last response.

Nine strip-filmed series of patterns were used. There were three different problem types involving form, size, or number, as the relevant dimension. A dimension was termed "relevant" when its use by *S* was requisite to identify a pattern with an appropriate key. Each problem type was further elaborated into one of three specific problems which differed in complexity, i.e., amount of irrelevant information. All irrelevant dimensions had a zero correlation with correct responses. When a dimension was neither relevant nor irrelevant, it appeared without variation at only one of its two levels within a given series.

In addition to the basic variations of relevant and irrelevant dimensions, some *S*s performed under a third condition wherein they received erroneous feedback. This feedback is interpreted by the *E* as MF. The probabilities that *S* would receive a "correct" signal over the 192 trials when he made a response actually correct with respect to the relevant dimension, were 1.00, .90, .80, .70, and .60 as determined by 0, 10, 20, 30, and 40% of MF, respectively. That is, on some occasions when *S* had made a "correct"

TABLE 1

Analysis of Variance on Errors

Source	*df*	*MS*	*F*
Misinformation feedback (MF)	3	452.58	50.45***
Linear	1	1,357.23	151.31***
Quadratic	1	.13	
Cubic	1	.40	
Complexity (C)	2	38.97	4.34*
Linear	1	76.26	8.50***
Quadratic	1	1.67	
Problems (P)	2	6.15	
Sex (S)	1	.55	
MF × C	6	56.58	6.31***
Linear	(3)	64.27	7.16***
Linear (1)	(1)	42.60	4.75*
Linear (2)	(1)	133.01	14.83***
Linear (3)	(1)	17.25	1.92
Quadratic	(3)	48.88	5.38**
Quadratic (1)	(1)	4.55	
Quadratic (2)	(1)	140.28	15.64***
Quadratic (3)	(1)	.01	
MF × P	6	6.66	
MF × S	3	7.68	
C × P	4	5.96	
C × S	2	11.57	1.29
P × S	2	4.32	
MF × C × P	12	16.01	1.78
MF × C × S	6	12.95	1.44
MF × P × S	6	12.67	1.41
C × P × S	4	16.09	1.79
MF × C × P × S	12	18.48	2.06*
Within	72	8.97	

* $P < .05$.
** $P < .01$.
*** $P < .001$.

response with respect to the relevant dimension, the signal lamp indicating a "wrong" response was lighted, deliberately misinforming *S* as to his choice of keys. Conversely, on some occasions, *S* received feedback to indicate he had responded correctly when, in fact, he had not done so. The scheduling of MF was determined by use of a table of random numbers. The order of pattern presentations was semirandom (i.e., all patterns appearing equally often but with the restriction that the same pattern was never presented on successive trials).

Design.—A 5 × 3 × 3 × 2 factorial design was used which incorporated the five degrees of MF (0, 10, 20, 30, or 40% of *S*'s responses met with erroneous feedback over 192 trials), 3 levels of complexity (1, 3, or 5 irrelevant dimensions), 3 problems (form, square, or triangle; size, large or small; or number, one object or two objects, as the relevant dimension), and 2 sex groups.

Results

The dependent variable studied was number of errors with respect to the *E*-defined relevant dimension over the 192 stimulus patterns. This score was analogous to the dependent variable used in guessing games and is not the number of errors as seen from *S*'s point of view.

Main effects.—The data were analyzed for the following four main effects: MF, Complexity, Problems, and Sex. The 192 stimuli were pooled into 12 blocks of 16 patterns each for the convenience of statistical and graphic treatment. Owing to the visible character of the graphs and the differences in learning rate, the analysis of variance was performed on the last two blocks (32 trials), omitting the Blocks variable and the 0% MF group. Summary of this analysis is presented in Table 1. Only the main effects of MF and Complexity were found to be significant. The increase of mean number of errors over the five degrees of MF, which was apparently linear, is shown in Fig. 1. Orthogonal polynomial analysis was performed on this function and only the linear component reached a significant level ($P < .001$). The significance of Complexity agrees with previous work on effects of this

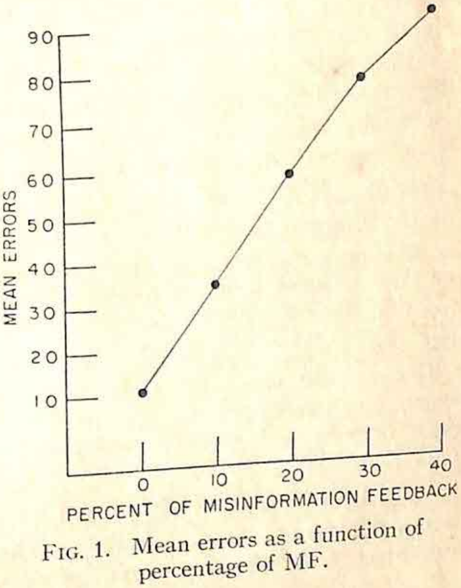


Fig. 1. Mean errors as a function of percentage of MF.

variable in concept identification (Archer, Bourne, & Brown, 1955; Bourne, 1957; Bourne & Pendleton, 1958; Brown & Archer, 1956). The mean numbers of errors for the conditions of 1, 3, and 5 irrelevant dimensions for the two blocks were 7.79, 8.73, and 10.31, respectively. The orthogonal polynomial analysis verified that only the linear term of Complexity reached significance ($P < .001$).

As expected, the main effect of Sex was insignificant. Neither was the Problems main effect significant, a result inconsistent with those of Archer, Bourne, and Brown (1955), and Brown and Archer (1956). The fact that the same irrelevant dimensions were used with the three relevant dimensions, supports the results of Grant and Curran (1952), and Heidbreder (1948, 1949), where identification of number and form was found to be equally difficult.

Following the suggestion of Lindquist (1953), graphic presentations of interactions were made. The interaction of MF × Complexity indicated that MF has a greater retarding influence on performance as task complexity increases. Further analysis of this interaction by orthogonal polynomials (Myers, 1959) revealed significant linear and quadratic trends. Both the linear and quadratic trends were broken down into their three components. It was shown that both main trends have significant quadratic components. In addition, the linear trend had a significant linear component (Table 1). This indicated that the slopes of MF × Complexity curves differ significantly on these components. Significance of the MF × C × P × S interaction is difficult to interpret.

An additional analysis of variance was performed on Blocks. This revealed a significant inverse relationship between number of trials and errors ($F = 41.62$, 11 and 990 df, $P < .01$). The interaction of Blocks × MF was also significant ($F = 4.29$, 44 and 990 df, $P < .01$), indicating that *S*s performing at the varying levels of MF approach different terminal levels of performance.

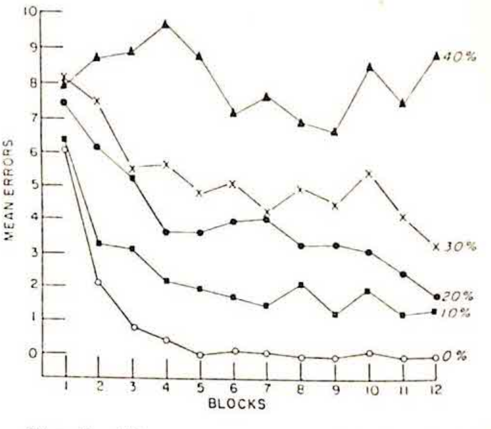


FIG. 2. Mean errors per block of 16 patterns. The parameter is percentage of MF. One irrelevant dimension.

Probability matching.—Figures 2, 3, and 4 represent mean number of errors per block for each MF group at each level of complexity. The obtained asymptotic value for each of the four degrees of MF was based on the mean number of errors on

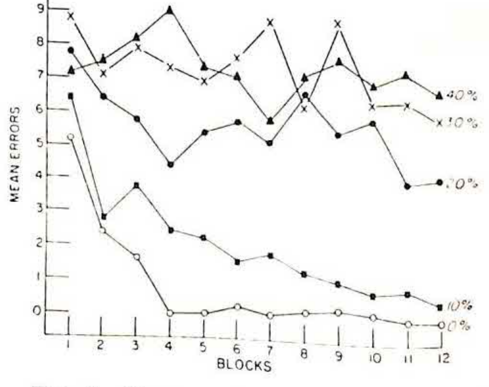


FIG. 3. Mean errors per block of 16 patterns. The parameter is percentage of MF. Three irrelevant dimensions.

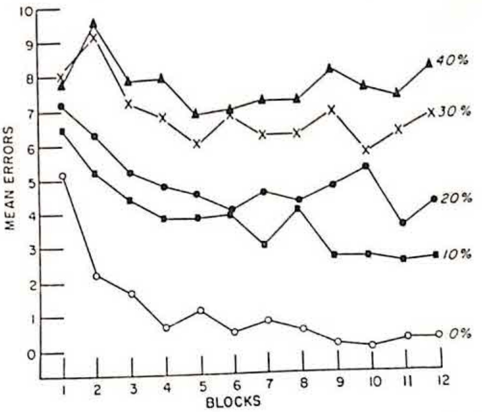


FIG. 4. Mean errors per block of 16 patterns. The parameter is percentage of MF. Five irrelevant dimensions.

Blocks 11 and 12, subgroups being pooled over the insignificant Sex and Problems variables. The expected error scores (probability matching) were determined by the probability of MF. Since there were 32 trials in the last two blocks, the expected error scores were 3.2, 6.4, 9.6, and 12.8 for 10, 20, 30, and 40% MF conditions, respectively. Differences were calculated between obtained and expected mean error scores. The hypothesis that the average difference scores over the three complexity groups would be zero was tested. Only the Mean source of variance for the 40% MF group was significant ($P < .01$). Only the 30% MF complexity groups differed significantly among themselves in the extent of their deviations from the predicted value ($P < .05$). This effect was linear as indicated by the orthogonal polynomial, with performance under greater degrees of complexity producing more errors than predicted by probability matching (Table 2).

A tabular analysis on number of left and right button responses revealed the same trend over trials throughout the various treatment conditions.

Restle's model.—In Restle's theory (1957), δ denotes the proportion of relevant cues in the problem. The values of δ for the three levels of complexity were estimated from the performance of the 0% MF groups. Using these δs, the value of cue validity, V, which would cause each of the 12 MF × Complexity groups to make the mean number of errors observed was computed by Equation 1. Following Restle's procedure (1957, p. 187), the values of V predicted by Equation 2 were found and compared with the computed Vs. The obtained V for each of the 12 conditions was subtracted from the predicted V, and an analysis of variance of these 12 difference scores was performed (Table

TABLE 2

ANALYSES OF VARIANCE FOR DEVIATIONS FROM EXPECTED MEAN ERRORS

Source	*df*	Misinformation Feedback							
		10%		20%		30%		40%	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Mean	1	12.02	.59	.59	.02	62.41	2.59	201.64	9.08**
Complexity	2	47.99	2.39	43.53	1.57	101.34	3.91*	35.09	1.58
Linear	1	60.36	3.01	60.17	2.20	170.67	6.58	20.16	.91
Quadratic	1	33.89	1.69	26.89	.98	30.68	1.18	50.00	2.25
Error	33	20.08		27.74		25.59		22.19	

* $P < .05$.
** $P < .01$.

TABLE 3

ANALYSIS OF VARIANCE FOR DIFFERENCES BETWEEN OBTAINED AND PREDICTED *V*s

Source	*df*	*MS*	*F*
Mean	1	.002	2.00
Cue Validity (*V*)	3	.061	61.00*
Linear	1	.18	180.00*
Quadratic	1	.00008	.08
Cubic	1	.0007	.70
Complexity (C)	2	.001	1.00
Linear	1	.0005	.50
Quadratic	1	.00104	1.04
$V \times$ C (Error)	6	.001	

* $P < .001$.

3). The main effect of *V* was significant ($P < .001$), indicating some significant discrepancy between predicted and obtained *V*s. This effect was linear as tested by the orthogonal polynomial; deviations becoming greater with increase in Complexity and MF.

One final theoretical analysis concerns the effect of number of irrelevant dimensions. Bourne and Restle (1959) assume that each such dimension is an equal source of irrelevant cues. From this they deduce that $1/\delta$ is a linear function of the number of irrelevant dimensions. Figure 5 is a plot of the reciprocal of estimated θ_T, which here should be proportional to $1/\delta$, in each of the three complexity conditions with all the MF conditions combined. The function is apparently straight-line in form.

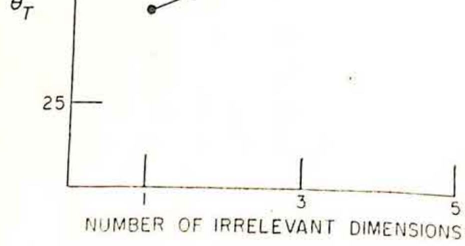


FIG. 5. $1/\theta_T$ as a function of number of irrelevant dimensions. Estimates based on the mean of total errors for each level of complexity over combined degrees of MF.

DISCUSSION

The results show that the efficiency of concept identification is a decreasing function of MF. This was expected from the results of Bourne and Pendleton (1958) in their study of probability of information feedback. In that study, the omission of feedback on a certain percentage of trials significantly reduced performance. This suggested strongly that MF would have a similar, though greater, effect. Consistent with previous work on effect of task complexity, the present investigation showed that irrelevant information has a quantifiable effect on task difficulty. Furthermore, the effect of MF depends on the level of task complexity, i.e., MF inhibits concept identification more as task complexity increases. It was shown that the slopes of MF × C curves differed significantly on both linear and quadratic components, producing decrement in performance.

The hypotheses concerning the effects of MF were based, in part, on the results of guessing game experiments. Owing to the complexity differences, rate of approach to asymptote ought to be less in the present task, but terminally, *S*s were expected to match probabilities. In the main, these expectations were realized. However, it is important to note that with greatest MF (40%), the number of errors was significantly greater than expected by probability matching. It is regrettable that more trials were not given, since the performance of certain groups might have surpassed probability matching performance, had further opportunity been provided. Trends of this sort were indicated in Fig. 2, 3, and 4.

In this study the sequence of MF was

randomly determined within the 192 patterns. In other studies, e.g., Morin's (1955) the percentage of MF was distributed over blocks of fewer trials, thus providing a more homogeneous distribution of MF. Further investigations of the effects of various distributions of MF would probably shed more light on the validity of V. It seems plausible to expect that a homogeneous distribution of MF would render it less effective as an inhibitor of performance. In relation to other studies of MF, present findings must be interpreted with caution. The *S*s' awareness of the existence or possibility of MF has a significant effect on performance as shown by Morin (1955). In this experiment, some indirect suggestion of the presence of MF was made in the instructions. It is probable that task difficulty could be increased by omitting this suggestion.

Some success was met in the theoretical analysis of data based on Restle's (1955, 1957) discrimination model. The θs, or rates of learning, were somewhat lower in the present data than is the case in simple guessing games (Restle, 1957). Furthermore, θ tends to decrease as task complexity increases, thus indicating that complexity accounts, at least in part, for the difference. The relationship between $1/\theta_T$ and number of irrelevant dimensions was linear as predicted. Quantitative predictions of errors (Equation 1), however, were grossly divergent from obtained errors for the four degrees of MF, and significant difference between estimated and predicted Vs were found. This result may stem from two sources. First, there were several unexpected inversions in the obtained error scores, i.e., trends in which errors did not consistently increase with combined increases in MF and Complexity (Fig. 2, 3, and 4). Such inconsistencies mediate against securing correct estimates of certain theoretical parameters, such as θ and V, since these parameters are based on the assumption of additivity of irrelevant cues. Second, it is possible that Restle's concept of cue validity is not appropriate in the analysis of problem solving behavior, since V is based on the function of π. It is important to note that π may be influenced by the differential distributions of MF. Bourne and Restle (1959) have provided evidence that the discrimination model can account for the results of some concept identification experiments, although the present study provides the first test of the cue validity formulation. Present results do not suggest any simple modification of this concept.

SUMMARY

The primary aim of this experiment was to investigate the effects of misinformative feedback (MF) and task complexity upon concept identification. Each of the 180 *S*s was given 12 blocks of 16 patterns, a total of 192 stimuli. The degrees of MF were 0, 10, 20, 30, and 40% and of complexity, 1, 3, and 5 irrelevant dimensions. There were three problems, each with a different relevant dimension, and equal numbers of men and women *S*s were included in each treatment combination. Essentially, *S*'s task was to learn the correct method for classifying visually presented geometric patterns.

The major findings were: (*a*) Performance improved significantly over trials. (*b*) MF was a significant source of variance in the analysis. The asymptotic mean error scores for the 10, 20, and 30% MF conditions did not significantly differ from the mean predicted values based on a probability matching hypothesis. (*c*) The relationship between errors and Complexity was shown to be linear. Analysis of variance produced a significant MF × Complexity interaction, suggesting an increase in the effect of MF at higher levels of Complexity. (*d*) Sex and Problems effects had no significant influence on performance. (*e*) The relationship between the reciprocal of the estimated proportion of relevant cues and the number of irrelevant dimensions was linear in accordance with Restle's (1955, 1957) learning model. (*f*) Number of errors as predicted by this model was greatly divergent from the obtained errors for all the combined MF × Complexity conditions.

REFERENCES

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, 49, 153–164.

BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

BOURNE, L. E., JR., & PENDLETON, R. B. Concept identification as a function of completeness and probability of information feedback. *J. exp. Psychol.*, 1958, **56**, 413–420.

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GRANT, D. A., & CURRAN, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HEIDBREDER, E. The attainment of concepts: VI. Exploratory experiments on conceptualization, at perceptual levels. *J. Psychol.*, 1948, **26**, 193–216.

HEIDBREDER, E. The attainment of concepts: VII. Conceptual achievements during card sorting. *J. Psychol.*, 1949, **27**, 3–39.

HUMPHREYS, L. G. Acquisition and extinction of a verbal conditioned response in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MORIN, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49**, 343–351.

MYERS, J. L. On the interaction of two scaled variables. *Psychol. Bull.*, 1959, **56**, 384–391.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.

(Received June 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

EYELID CONDITIONING PERFORMANCE UNDER PARTIAL REINFORCEMENT AS A FUNCTION OF UCS INTENSITY[1]

LEONARD E. ROSS[2] AND KENNETH W. SPENCE

State University of Iowa

Investigations of the relationship of UCS intensity to performance in eyelid conditioning have consistently employed continuous (100%) reinforcement. Thus, while studies have demonstrated (Spence, 1958) that performance is an increasing function of the intensity of the UCS, evidence is not available concerning the relationship of this variable to eyelid conditioning performance under partial reinforcement schedules. Although there are a number of eyelid conditioning studies in the literature that have used 50% partial reinforcement (Grant & Hake, 1951; Grant & Schipper, 1952; Grant, Schipper, & Ross, 1952; Gynther, 1957; Ross, 1959), a comparison of the performance levels obtained in these experiments as a function of the UCS strength employed is not feasible since in some studies the puff strength was unspecified and, in addition, experimental conditions, which might in themselves account for performance differences, have varied from study to study.

Under 100% reinforcement conditions the relationship between conditioning performance (percentage of CRs, Trials 41–80) and the intensity of the UCS has been found to be a negatively accelerated increasing function (Spence, 1958). The question arises whether or not the relationship under partial reinforcement conditions is of the same general form. The previously cited studies which employed partial reinforcement uniformly found higher levels of conditioning with 100% reinforcement than with 50% partial reinforcement. Since this continuous reinforcement superiority is maintained when comparisons are made in terms of reinforced trials rather than trials alone, the lower level of performance of the partial groups is presumably due to the decremental effects of nonreinforced trials. It has been suggested (Ross, 1959) that these effects are, in part at least, due to factors that might be classified as "set" variables (i.e., under some degree of voluntary control).

Whatever the nature of the factors producing the decrement it seems quite possible that the introduction of nonreinforced trials might result in a relationship between conditioning level and UCS intensity different from that obtained with continuous reinforcement. In the present experiment groups of *S*s under a random 50% reinforcement schedule were conditioned with different intensities of the UCS in an effort to provide information concerning the function relating conditioning performance under partial reinforcement conditions to the strength of the UCS. In addition, one group was conditioned with continuous reinforcement to supplement existing 100% data so that 50%–100% comparisons could be made at an additional UCS intensity value.

---

[1] This study is part of a project concerned with the influence of motivation on performance in learning under Contract N9 onr-93802, Project NR 154-107 between the State University of Iowa and the Office of Naval Research. Robert S. Blanchard helped in the collecting of the data.

[2] Now at the University of Wisconsin.

METHOD

Subjects.—The *S*s were 115 women and 58 men from an introductory psychology course. Of these, 10 men and 6 women were discarded for exceeding the criterion used to exclude voluntary responders (Spence & Ross, 1959), while 1 woman was excluded from the data for giving CRs to test trial presentations of the CS alone. Six additional *S*s were discarded, three due to *E* error and three due to apparatus failure. The remaining 150 *S*s were assigned at random to five groups except that the proportion of women to men was fixed at a ratio of 7 to 3.

Apparatus and method of recording.—The apparatus and method of recording which involved a potentiometer–polygraph system, has been described elsewhere (Spence & Taylor, 1951). The CS was an increase in brightness of a $2\frac{1}{4}$-in. diameter circular milk glass disk from less than .004 apparent ft-c to .32 apparent ft-c. The UCS was a 50-msec. air puff delivered through a .062-in. diameter orifice to the right eye.

Conditioning procedure.—Each *S* was given instructions to blink once to the ready signal and then to look at the disk on the wall, remaining as relaxed as possible, until the light went out. Following the reading of the instructions three test trial presentations of the CS alone and one presentation of the UCS alone were given. Immediately following this preliminary procedure each *S* received 140 trials at intertrial intervals of 15, 20, and 25 sec., averaging 20 sec., and arranged according to a fixed schedule. The word "ready" preceded each presentation of the CS by 2, 3, or 4 sec. according to a prearranged schedule.

Four groups of 30 *S*s were conditioned with a 50% reinforcement schedule consisting of half reinforced and half nonreinforced trials, presented in a prearranged sequence within which neither type of trial occurred more than twice in succession. A reinforced trial involved a 500-msec. CS-UCS interval while on nonreinforced trials the UCS followed the CS onset at 2500 msec. Each group was conditioned with a different intensity of UCS. The four values used were .33, .6, 1.0, and 4.0 psi. An additional group of 30 *S*s was given 100 trials with continuous reinforcement. The UCS for this group was a .33 psi air puff.

RESULTS AND DISCUSSION

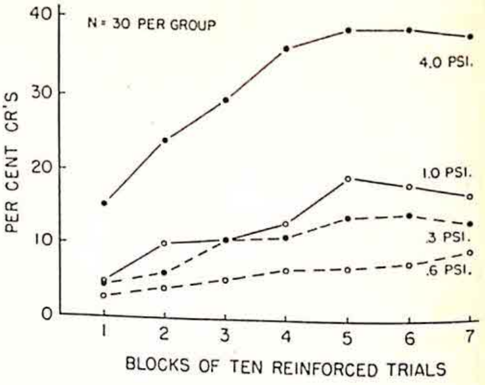


FIG. 1. Mean percentage of CRs in successive blocks of 10 reinforced trials.

Frequency curves of conditioning by blocks of 10 reinforced trials for the four groups receiving partial reinforcement are shown in Fig. 1. Inspection of the figure reveals that only the 4.0-lb. psi group showed an appreciable performance increase, the other three groups remaining under the 20% CR level throughout the course of conditioning. Although the 1.0-lb. group was superior to the .33-lb. group the performance level of the .6-lb. group was slightly lower than that of the .33-lb. group.

To statistically evaluate these differences a Kruskal-Wallis (1952) *H* test was employed with the scores for the four groups over conditioning Trials 51–70 (Trials 100–140). The resulting *H* of 20.64 was significant at the .001 level. Subsequently, Mann-Whitney *U* tests were used to test the observed differences for

TABLE 1

MANN-WHITNEY *U* AND THE CORRESPONDING *P* VALUES FOR ALL POSSIBLE COMPARISONS OF THE GROUPS ON TRIALS 41–70

Group Comparisons	*U*	*P*
.33– .6	.75	.453
.33–1.0	.90	.368
.33–4.0	3.47	.0006
.6 –1.0	1.65	.099
.6 –4.0	4.09	.0001
1.0– 4.0	2.80	.005

all possible pairs of puff strength values. Table 1 presents the *U* and corresponding *P* value for each of the comparisons. It should be noted that while the 4.0-lb. group differed significantly from each of the other groups, none of the other comparisons reached the .05 level of significance. In view of these statistical results the differences among the three low puff strength groups must be considered to be fortuitous. It would appear that with UCS values of 1.0 lb. and less a low performance level is obtained and that within this range there is little, if any, positive relationship between UCS intensity and conditioning performance.

In order to compare the function relating conditioning performance to UCS intensity for 50% partial and continuous reinforcement conditions, the data of a number of published studies from the Iowa laboratory were used. Except for the intensity of the UCS these studies were highly comparable to the present study in terms of apparatus, conditioning procedure, and experimental conditions (i.e., CS, CS-UCS interval, intertrial interval, etc.). Since the use of a frequency measure imposes a ceiling on performance (100%), and a number of *S*s in the strong puff groups had reached this ceiling, the comparisons were made in terms of the conditioning performance of homogeneous *S*s, i.e., the middle 40% of each group. Thus the *S*s from each group were rank ordered on the basis of the total number of responses made during the first 70 conditioning trials and the mean percentage of CRs made by the middle 40% of each distribution was computed. Because of varying ratios of men to women this procedure was followed for each sex separately and the mean percentage values obtained were averaged for each experiment. The results obtained are presented in Fig. 2. For the continuous reinforcement function the .33-lb. point was taken from the 100% reinforcement group in the present experiment while the other points, from .6 to 4.0 lb., respectively, were obtained from studies by Runquist and Ross (1958), Runquist and Spence (1959), Ross (1959), and Spence and Ross (1959). All of the points on the partial reinforcement function were taken from the present experiment except for the 2.0-lb. point which was obtained from an earlier study (Ross, 1959).

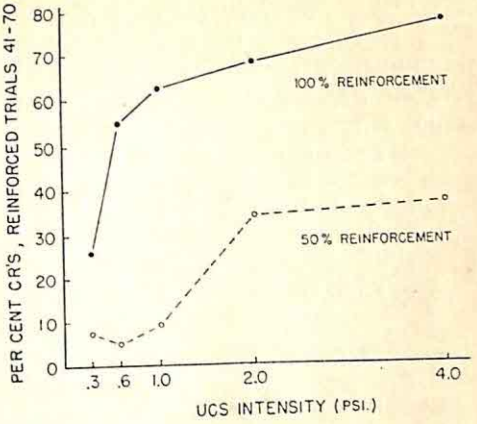


FIG 2. Relationship between percentage of CRs and UCS intensity for groups of homogeneous *S*s (see text) receiving partial and continuous reinforcement.

From an inspection of Fig. 2 it is evident that the shapes of the functions are quite different for the two reinforcement schedules. With continuous reinforcement the function is negatively accelerated with small increases in puff strength at low absolute levels resulting in large increases in performance. In contrast, under 50% reinforcement the function is apparently S-shaped with little performance change over the low range of UCS intensity values.

Just why such different results should be obtained with the introduction of nonreinforced trials into the conditioning situation is not clear. However, it is possible that, if the decrement found with 50% reinforcement is to some degree related to "set" factors, inhibition of performance is more readily accomplished under conditions of low puff strengths. Thus, increases in puff strengths over this range would be ineffective in increasing performance. The large differences between the 100% and 50% reinforcement groups at high levels of puff strength require that considerable "inhibition" still be present with such puffs. Stronger puffs, then, would increase performance but not completely overcome the "inhibition" effects of 50% reinforcement conditions. Further clarification of the differences between the functions would appear to be dependent upon added information concerning the role of inhibitory factors in partial reinforcement.

SUMMARY

Groups of *S*s under a random 50% reinforcement schedule were conditioned with different intensities of UCS (.33, .6, 1.0, 4.0 psi) in an effort to provide information concerning the function relating conditioning performance under partial reinforcement conditions to the strength of the UCS. At UCS values of 1.0 lb. and less, a low performance level was obtained, with little indication of a positive relationship between performance and UCS intensity. A 4.0-lb. UCS resulted in performance significantly higher than that of the other groups.

In order to compare the function relating conditioning to UCS strength obtained for partial reinforcement in the present experiment with one for continuous reinforcement, the data from a number of previously published studies were analyzed and a similar function determined for 100% reinforcement. The functions differed in shape, the partial reinforcement function being S-shaped, while that for continuous reinforcement appeared to be negatively accelerated throughout.

REFERENCES

GRANT, D. A., & HAKE, H. W. Dark adaptation and the Humphries random reinforcement phenomenon in human eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 417–423.

GRANT, D. A., & SCHIPPER, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

GRANT, D. A., SCHIPPER, L. M., & ROSS, B. M. Effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

GYNTHER, M. D. Differential eyelid conditioning as a function of stimulus similarity and strength of response to the CS. *J. exp. Psychol.*, 1957, **53**, 408–416.

KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. Stat. Ass.*, 1952, **47**, 583–621.

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

RUNQUIST, W. N., & ROSS, L. E. A preliminary investigation of the relationship between GSR, heart rate change, and eyelid conditioning. *Proc. Iowa Acad. Sci.*, 1958, **65**, 398–404.

RUNQUIST, W. N., & SPENCE, K. W. Performance in eyelid conditioning as a function of UCS duration. *J. exp. Psychol.*, 1959, **57**, 249–252.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

(Received June 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

EFFECT OF TRIPLET AND QUADRUPLICATE LOCATION IN VERBAL MAZE PATTERNS UPON SERIAL POSITION ERRORS[1]

GEDIMINAS NAMIKAS, CHARLES P. THOMPSON, AND W. J. BROGDEN

University of Wisconsin

The effect of change in pattern of the 16-unit, 4-choice verbal maze of Brogden and Schmidt (1954) upon its acquisition and form of the serial position error curve has been studied by Thompson, Voss, and Brogden (1957) and by Ernst, Hoffeld, Seidenstein, and Brogden (1960). Alteration of the standard maze pattern by the introduction of a doublet (two successive like response items) or a split-doublet (two identical response items separated by a different response item) changes the form of the serial position error curve but does not affect speed of acquisition. A doublet in each of five different locations of the maze pattern produces an increase in error at the position of the second item. A split-doublet in each of four different locations (except for Positions 2–4) produces an increase in error at the third position (second like item). With the one exception for the split-doublet, there is no differential effect of doublet or split-doublet location in the maze pattern.

The doublet effect is maximum for the total error measure early in acquisition and decreases in magnitude thereafter. For the first error measure, there is no doublet effect over all trials. For the first one-quarter of trials the doublet effect is present as an increase of error at the position of the second item, but for the last three-quarters of trials the effect is reversed and occurs as a decrease in error at the second position of the doublet. The presence of a doublet in the maze pattern thus has an inhibitory effect upon acquisition during the discovery phase and this is a function primarily of the guessing habits of *S*s, particularly the tendency not to repeat a response correct at the previous position. Once discovery is completed, the identity of the doublet items facilitates further acquisition of the second item of the doublet.

The split-doublet phenomenon is apparently different from that for the doublet. There is no evidence of difference between the total error and first error measures over all trials or for stages of acquisition, nor of a relation between guessing habits of *S*s and the split-doublet effect. These results suggest that there may be no consistent effect of a pattern dimension upon the serial position error curve and that variables representative of the dimension such as the doublet and split-doublet have specific and independent effects. Within the parameters of the 16-unit, 4-choice verbal maze pattern, with each choice represented four times, the triplet (three successive like responses) and the quadruplicate (four successive like responses) are the only other sources of pattern variation that can be included in a dimension that also includes the doublet and split-doublet. The effects of the triplet

---

[1] This research was supported in part by grants from the National Science Foundation and the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

and quadruplicate upon the form of the serial position error curve were hypothesized to be different from those of the doublet and split-doublet, to be different from each other, to involve both increases and decreases in error as a function of position within the triplet or quadruplicate sequence, and to show no difference in character as a function of location within the maze pattern. It was hypothesized further that the introduction of a triplet or quadruplicate would increase the speed of acquisition over that for the standard maze pattern. This paper presents the results of two experiments, one in which the triplet and its location represent the independent variable and the other for which the independent variable is the quadruplicate and its location in the maze pattern.[2] These two experiments follow closely the design of the earlier studies on the doublet and split-doublet in an attempt to test the hypotheses presented above.

PROCEDURE

Experiment I.—The four experimental groups acquired a modification of the standard maze pattern which includes a single triplet located at Maze Units 3, 4, and 5, at 6, 7, and 8, at 9, 10, and 11, and at 12, 13, and 14, respectively. Each modified pattern was altered not only to locate the triplet at the appropriate positions, but also to exclude doublet, split-doublet, or other noticeable sequences in the maze. Each pattern was permuted to obtain four subpatterns and thus provide for each number (10, 20, 30, or 40) to be represented once at each of the 16 positions. Data for the control pattern is that reported by Ernst et al. (1960).

Of 20 *S*s in each of the four experimental groups, five learned each one of the four subpatterns. All of the *S*s were treated by a single *E*. The *S*s were volunteer students from the classes in elementary psychology at the University of Wisconsin. Assignment of *S*s to experimental conditions was random within each replication.

The instructions of Thompson et al. (1957) were read to *S*. Responses of 10, 20, 30, or 40 were made by *S* with *E* signaling the correct response by saying "Forward." This procedure continued until *S* completed one errorless trial. Each response at each unit was recorded by *E* for each trial. A stop watch was used to measure the time of each trial and the 1-min. rest between trials. Other conditions were comparable to those of Thompson et al. (1957).

Experiment II.—The patterns learned by the three experimental groups contained a single quadruplicate located at Maze Units 3, 4, 5, and 6, at 7, 8, 9, and 10, and at 11, 12, 13, and 14, respectively. Each pattern represents a limited modification of the standard pattern to provide a quadruplicate at the desired locus without other noticeable pattern sequences elsewhere in the maze. In all other respects, the procedure for Exp. II was identical to that of Exp. I.

RESULTS

Experiment I.—Analyses of variance of the data for time and trials for the four experimental and one control group showed no significant effect for maze pattern, maze pattern subgroups, or the interaction of these two variables. Since the assumption of homogeneity of variance was found not tenable for the error data, the 1% level of significance was selected in lieu of the 5% level for all analyses. No source from the between-*S*s variation was significant. Further analysis of the significant interaction of serial position by maze pattern was based upon several conversions of the data.

Total error to the criterional trial at each serial position was converted to percentage of total error for each *S* in each group (McCrary & Hunter, 1953). The group mean of the percentage scores for *S*s of each experimental group and the combined control group of Ernst et al. (1960) are presented in Fig. 1 by

---

[2] Thompson collaborated with the senior author on the triplet experiment and Namikas on the experiment with the quadruplicate.

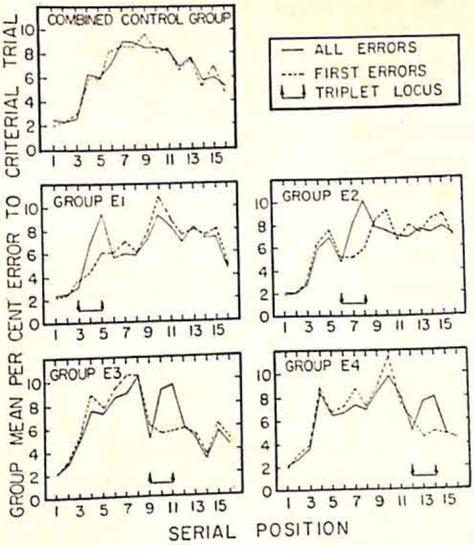


FIG. 1. Serial position curves in terms of percentage of total errors and percentage of first errors as a function of triplet locus.

the solid line curves. There are substantial differences in the form of the curves for the experimental groups compared with each other and with the control group. These differences are clearly shown by the solid line curves in the left half of Fig. 2. The data for these curves were obtained by taking the difference in percentage of error for each *S* and the mean percentage error for the combined control group at each of the 16 serial positions and computing the mean of these differences for each of the four experimental groups. For each group there is an increase in error for the third position of the triplet, and except for Group E_1, there is a decrease in error for the first position. Analysis of variance of these data shows in each case a significant interaction of serial position and maze pattern and thus establishes significant differences in the form of the difference curves between the experimental groups and the control group and between the four experimental groups.

The broken line curves of Fig. 1 are the group mean percentage of first errors over all trials to the criteria and the broken line curves of Fig. 2 are the difference curves between each experimental group and the control group. The curves in the right half of Fig. 2 give the difference between percentage total error and percentage first error for the experimental groups. The only substantial differences between the all error and first error measures occur at Positions 2 and 3 of the triplet. Analyses of variance of the two sets of first error data and of the differences between percentage total error and percentage first error show in each case a significant serial position by maze pattern interaction.

The *t* test was used to establish fiducial limits at the 1% level of significance for the three sets of mean difference data shown in Fig. 2.

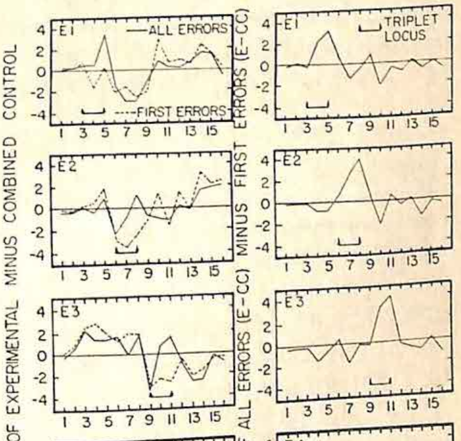


FIG. 2. Curves of differences between serial position curves for experimental minus control groups for percentage of total and first errors (left side) and curves of differences between these differences for each experimental group (right side).

These limits are ±3.1 for all errors, ±2.9 for first errors, and ±1.6 for the difference between all errors and first errors. For the all error measure, the differences from the control group that equal or exceed the fiducial limits for triplet locus are the first triplet item (negative) for Group E_3, and the third triplet item (positive) for Group E_1. No positions outside of the triplet locus exceed the limits.

For the first error measure, the first item of the triplet exceeds the negative limit for Groups E_2 and E_3, and the second item of the triplet exceeds the negative limit for Groups E_2 and E_4. Other significant differences are Positions 10 (positive) for E_1, 14 (positive) for E_2, 4 (positive) for E_3, and 10 (positive) for E_4.

For the difference between total error and first error, the second and third items of the triplet exceed the positive fiducial limit for every group. In addition, Position 10 (negative) exceeds the limits for E_1, 10 (negative) and 14 (positive) for E_2, 7 (negative) for E_3, and 7 and 10 (negative) for E_4.

The triplet effect is not entirely consistent for all groups and is of different character for the total and first error measures. It appears to be one of increasing total error over the three positions, starting with less error than that for the control pattern at the first position and terminating with greater error. The means over all groups are −1.8, 0.0, and 2.1, respectively, for the three positions. For the first error measure, the effect seems to be a general reduction in error from that of the control group at all positions, that increases from the first to second item and decreases from the second to the third. The means over all groups are −1.5, −2.6, and −0.8, respectively. These apparent trends were tested by orthoganol polynomial analysis of variance (Grant, 1956) on the data at the triplet positions over all experimental groups for total error and for first errors. A significant linear trend was found for the total error measure and a significant quadratic trend occurred for the first error measure.

Since Ernst et al. (1960) found the difference in doublet effect for the total and first error measures to be a function of stage of acquisition, the frequency of total and first error at the triplet locations were tabulated by quarters of trials to the criterional trial for the experimental groups and the control group. The E-CC differences were obtained for each *S* of the experimental groups and the over-all group means for total and first error measures computed for each triplet position. All differences for both measures are positive for the first quarter of trials and negative for each one of the last three quarters of trials except for the total error measure for the third triplet item for the second quarter. The linear trend for total errors and the quadratic trend for first errors do not occur consistently for each quarter of trials, but for both measures the effect appears to be maximum for the first quarter and to decline progressively over the remaining three quarters. Each measure shows its characteristic trend, but early in learning the trend occurs over errors greater than those for the control group and later in acquisition, over errors less than the control group. No statistical tests were made on these data.

Experiment II.—Analyses of variance of the data for time and trials for the three experimental and one control group showed no significant effect for maze pattern, maze pattern subgroups or the interaction. Since the assumption of homogeneity of variance was untenable for the error data, the 1% level of significance was selected in lieu of the 5% level for all analyses. Treatment of the error data for this experiment follows that described already for Exp. I, with comparable results. The interaction of Serial position × Maze pattern is significant for total errors, for percentage total errors, for percentage first errors, for percentage total errors experimental minus control, and for percentage first errors experimental minus control. The serial position error curves for percentage of all error and for percentage of first error for the control group and the three experimental groups are presented in Fig. 3. The effect of the quadruplicate appears to be one of general reduction of error. This is shown by the curves of difference between the experimental and control groups for percentage total errors and percentage first errors which are presented in the left half of Fig. 4. The curves in the right half of Fig. 4 are the differences between percentage total errors and percentage first errors for each experimental group. The curves for all error and for first error follow each other closely except that the first error curves for E_1 and E_2 appear to show a greater reduction in error for the quadruplicate locus than do those for total error. The curves of difference between total and first errors appear to reflect these differences primarily. Analysis of variance of the difference data between total and first errors reveals a significant interaction of serial position and maze pattern.

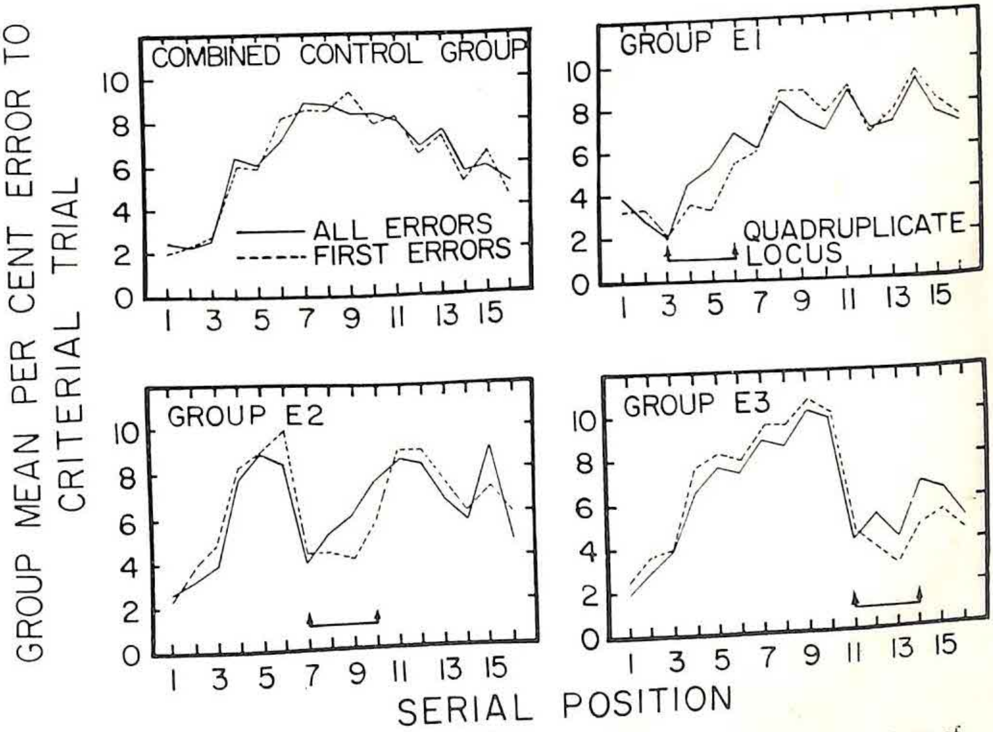


FIG. 3. Serial position curves in terms of percentage of total errors and percentage of first errors as a function of quadruplicate locus.

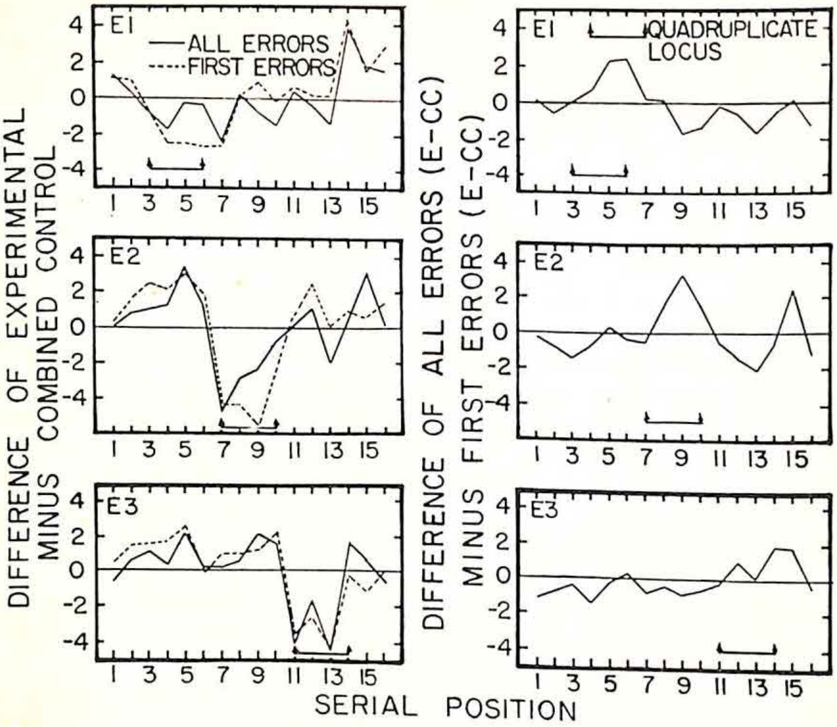


FIG. 4. Curves of differences between serial position curves for experimental minus control groups for percentage of total and first errors (left side) and curves of differences between these differences for each experimental group (right side).

The fiducial limits at the 1% level of significance are ±3.6 for all errors, ±2.0 for first errors, and ±2.6 for the difference between all errors and first errors. For the total error measure, the differences from the control group that equal or exceed the negative limit at the quadruplicate locus are Item 1 for E_2, and Items 1 and 3 for E_3. The only other position that exceeds the limits is 14 (positive) for E_1.

For the first error measure, quadruplicate Items 2, 3, and 4 for E_1, 1, 2, 3, and 4 for E_2, and 1, 2, and 3 for E_3 equal or exceed the negative limit. Other positions that exceed the limits are 7 (negative), 14, and 16 (positive) for E_1, 3, 4, 5, and 12 (positive) for E_2, and 5 and 10 (positive) for E_3. The fiducial limits for the curves of difference between all errors and first errors are exceeded by Item 3 (positive) of the quadruplicate for E_2 and by Position 15 (positive) for the same group.

Orthogonal polynomial analyses of variance (Grant, 1956) were completed on the E-CC data of the quadruplicate positions over all experimental groups for total errors and for first errors. The over-all group means are −3.3, −2.1, −2.3, and 0.1, respectively, for total errors, and −1.0, −0.7, −0.5 and 0.5, respectively, for first errors. Significant linear, quadratic, and cubic components were obtained for the total error data, and significant quadratic and cubic components were obtained for the first error data.

The E-CC data for total and first errors at the positions of the quadruplicate were each tabulated by quarters of trials to the criterial trial. All means over groups show less error than the control group except the last three positions of the quadruplicate for the first quarter of trials for the total error measure. These latter means are progressively positive in magnitude. The trend over the four positions of the quadruplicate is more apparent during the first two quarters of trials for both measures and is more evident with the total error measure than with the first error measure. No statistical tests were made of these data.

DISCUSSION

The two experiments of this study establish a triplet effect and a quadruplicate effect on the form of the serial position error curve. Locus of the triplet and the quadruplicate does not appear to be differential, except that location of the pattern early in the maze (3, 4, 5 for the triplet and 3, 4, 5, 6 for the quadruplicate) produces a less clear-cut effect than does location of either pattern elsewhere. The general form of each effect is different for the total error and first error measures. For the triplet, there is a linear trend for total error, starting with a decrease in error relative to the control pattern for the first item, no difference for the second item, and an increase in error for the final item. There is a quadratic trend for the first error measure, with less error than the control pattern for each of the triplet items. For the quadruplicate, both the total error and first error measures show less error than the control pattern for the first three positions and a zero or small positive difference for the fourth item. The general trend for the total error measure has linear, quadratic, and cubic components and that for the first error measure has quadratic and cubic components. For both the triplet and quadruplicate, there are other differences in form of the serial position error curve that may not be accounted for by the location of either pattern or by the interchanges of response items necessary to pattern formation.

The triplet and quadruplicate effects are not only of different character from each other but differ also from the doublet and the split-doublet effects (Ernst et al., 1960). The doublet effect stands apart from the other three pattern effects in that it is in part a function of the guessing habits of *S*s and the effect measured by first errors is diphasic. Although the quadruplicate effect as measured by total errors shifts from positive error relative to the control group for the first quarter of trials to negative for each of the last quarters, the trend over both triplet and quadruplicate positions is relatively stable over the course of acquisition. The differences in the triplet and quadruplicate effects between total and first errors appear to be a function of the differences in the nature of these measures.

It does appear that as the number of like items is increased, there is a progressive decrease in error at the locus involved, going from an initial increase in error relative to the control pattern for the doublet to a decrease in error maximum for the quadruplicate. All show a trend within the pattern of a relative increase in error from the initial to succeeding items. Each represents a characteristic modification in the form of the serial position error curve without affecting the over-all measures of acquisition.

The quadruplicate represents the upper limit of a pattern dimension involving succeeding like items with fixed parameters of a 16-unit maze of four choices, each of which occurs as the correct response four times. Longer mazes with a greater number of choices would permit extension of the variable of number of succeeding like items and should produce larger decrements in error relative to the control pattern than that produced by the quadruplicate. It is likely also that the pattern effects beyond the quadruplicate in longer mazes will produce an over-all decrease in measures of acquisition relative to the control pattern and differential effects as a function of locus. The results from the triplet and quadruplicate studies suggest that their effects are reduced for the early positions of the maze, as is the case also for the split-doublet.

SUMMARY

Two experiments were designed to test respectively the effect of location of a triplet and a quadruplicate on the form of the serial position error curve from the acquisition of a 4-choice, 16-unit verbal maze. Significant differences were obtained in both experiments in the form of the curves between each experimental group and the control group, and between experimental groups. The effect was less clear for the locus early in the pattern (3, 4, 5 for triplet and 3, 4, 5, 6 for quadruplicate) than for the other locations. The triplet effect is a linear trend over the three positions for the total error measure, with over-all mean differences relative to the control group of −1.8, 0.0, and 2.1. For the first error measure, the respective means are −1.5, −2.6, and −0.8, with a significant quadratic trend. The quadruplicate effect is represented by mean differences over the four positions of −3.3, −2.1, −2.3, and 0.1 for total errors and −1.0, −0.7, −0.5, and 0.5 for first errors. For the former, there are linear, quadratic, and cubic components, and for the latter, quadratic and cubic components. No relation was found between these phenomena and the guessing habits of *S*s. The phenomena are relatively stable over successive stages of acquisition, and there are no marked differences between total and first errors at the different stages of acquisition. The results are discussed relative to the doublet and split-doublet effects and in relation to a general pattern dimension and its effect upon the serial position error curve.

REFERENCES

BROGDEN, W. J., & SCHMIDT, R. E. Effect of number of choices per unit of a verbal maze on learning and serial position errors. *J. exp. Psychol.*, 1954, **47**, 235–240.

ERNST, R. L., HOFFELD, D. R., SEIDENSTEIN, S., & BROGDEN, W. J. Relation of serial position errors to doublet and split-doublet location in verbal maze pattern. *J. exp. Psychol.*, 1960, **59**, 94–103.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

MCCRARY, J. W., JR., & HUNTER, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131–134.

THOMPSON, R. F., VOSS, J. F., & BROGDEN, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, 1957, **54**, 253–258.

(Received June 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

EFFECT OF CONCURRENT RESPONSES ON THE EVOCATION AND GENERALIZATION OF THE CONDITIONED EYEBLINK[1]

G. ROBERT GRICE[2] AND JOHN D. DAVIS[3]

University of Illinois

Recently, the authors (Grice & Davis, 1958) have presented evidence for a mediated generalization effect in human eyelid conditioning. The experiments involved a differential conditioning paradigm in which the middle of three tones served as the CS. The *S*s also made concurrent manual reaction-time responses to the three tones. It was found that significantly more CRs were made to the negative stimulus to which the manual response was the same as the CS than to the one to which the manual response was different. This asymmetrical generalization gradient was interpreted as being in accord with a theory of mediated stimulus generalization, and the advantage was claimed for the design that the "mediating reactions" were overt and subject to experimental control. The results of these experiments were seriously complicated, however, by a direct facilitative effect of the manual reactions on the eyeblink response. This effect appeared to be more powerful than the mediational one and introduced some danger in the interpretation of the experiments. For example, if the manual reactions to the "Same Response" tone should be faster than to the "Different Response" tone the asymmetrical generalization could be an artifactual result of differential facilitation, rather than an associative, mediational effect.

The first purpose of the present experiments is to investigate the relation between these two effects by means of an analysis of the relation between the latencies of the manual reactions and the occurrence of eyeblinks. The second purpose is to compare the mediational effect in a situation where the CS and the generalized stimuli are in the same sense modality with one in which they are different. Theoretically, while primary stimulus generalization would be substantially greater within one modality, the amount of mediated generalization should be independent of the fact that the CS and generalized stimuli are in different sensory dimensions.

METHOD

Apparatus

The experimental situation was the same as previously described (Grice & Davis, 1958) with the exception of certain improvements in instrumentation. The tones were presented by means of Permoflux PDR-8 earphones rather than by means of a loudspeaker. The same three tones, 240, 850, and 1900 cps, were presented at 70 db loudness level with reference to an average threshold for 1000 cps. They were switched on and off by means of a Grayson-Stadler electronic switch, which was set with a 30 msec. rise and decay time. A continuous "white" masking noise of 20 db sensation level was also presented. The UCS was a puff of air of 1 lb./sq. in. pressure, and the CS-UCS interval was .5 sec. Programming equipment made possible the automatic presentation of

---

[1] This research was supported by a grant from the National Science Foundation.

[2] In the absence of the junior author, the senior author was entirely responsible for the preparation of the present report.

[3] Now at the American University of Beirut.

the entire experimental session of 100 trials, with attendant reduction of *E* error and general increase in precision.

The manual reaction latencies were recorded on the oscillograph tape with the CRs and were read to the nearest 8 msec. Reactions in excess of 640 msec. were recorded at that value.

Procedure

Experiment I.—With the exception of the apparatus differences noted, the procedure was the same as in Exp. I of the previous paper. The same counterbalancing of the high and low tones and right and left hand responses was employed. The middle tone always served as the CS, the air puff never being paired with either of the other tones. Criteria for scoring CRs and elimination of voluntary responders were the same, although no voluntary responders were actually eliminated in either experiment.

The *S*s were 12 men and 12 women students who served as part of a requirement for a course in general psychology. Eight additional *S*s who made fewer than 12 CRs were eliminated. The number of errors in the reaction-time task ranged from 0 to 6 with a mean of 2.0 and a median of 1.85.

Experiment II.—Procedure was the same as in Exp. I, except that the CS was a frosted 7-w. bulb placed $\frac{3}{4}$ in. behind the 1-in. milk glass disk used as a fixation point in the previous experiments. The negative stimuli were the same 240- and 1900-cps tones used in Exp. I.

The *S*s were 12 men and 12 women from the same population as Exp. I. Three *S*s who made fewer than 12 CRs were eliminated. Errors on the reaction-time task ranged from 0 to 5, with a mean of 2.21 and a median of 2.0.

RESULTS

All data presented are for Trials 41–100, which included 20 presentations of each of the three stimuli. The mean number of CRs to each of the three stimuli in the two experiments are presented in Table 1. The predicted asymmetry was obtained in both experiments, there being more responses to the negative stimulus to which the manual reaction was the same as to the CS than to the one for which it was different. Experiment I replicates the results of Exp. I in the previous paper except that the above-100% generalization to the Same Response stimulus found in that experiment was not obtained.

Mean latencies of the manual responses are presented in Table 2. For Exp. I, the reaction times were somewhat faster to the Same Response stimulus. There was little difference between the reaction times to these two stimuli in Exp. II.

In order to more fully evaluate the effects of the mediational and facilitative effects, point biserial correlation coefficients were computed over trials between reaction time and the occurrence or nonoccurrence of CRs for each *S* separately. The means of these correlations, based on the *z* transformation, are presented in Table 2. The number of CRs to each stimu-

TABLE 1

MEANS OF UNADJUSTED AND ADJUSTED NUMBER OF CRs

Stimulus	Experiment I				Experiment II			
	Unadjusted		Adjusted		Unadjusted		Adjusted	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Same Resp.	13.25	4.18	13.01	3.69	8.33	5.56	8.91	5.26
CS	14.46	3.55	14.15	4.08	16.00	2.79	14.73	3.06
Diff. Resp.	11.00	5.77	11.56	5.10	5.96	6.19	6.66	5.87

TABLE 2

REACTION-TIME MEANS AND AVERAGE POINT BISERIAL CORRELATIONS BETWEEN REACTION TIME AND CR OCCURRENCE

Stimulus	Reaction Time in Msec.				r_{pbis}			
	Exp. I		Exp. II		Exp. I		Exp. II	
	Mean[a]	*SD*	Mean[a]	*SD*	Mean[b]	Range	Mean[b]	Range
Same Resp.	508	73	500	59	−.63	+.30 to −.89	−.58	−.11 to −.86
CS	509	78	454	36	−.59	−.08 to −.81	−.63	−.36 to −.87
Diff. Resp.	526	61	503	64	−.46	+.04 to −.86	−.29	+.32 to −.78

[a] Mean of *S* means.
[b] Obtained by *z* transformation.

lus was then adjusted for each *S* separately by means of his own "within stimulus" regression coefficient. This leaves the *S* totals unchanged, but adjusts the number of CRs to each stimulus for each *S* for reaction time, and, presumably, for within-*S* variation in the facilitative effect. Means of the adjusted measures are presented in Table 1. It may be noted that the asymmetry of generalization was preserved for both experiments.

Separate analyses of variance of the adjusted numbers of CRs for the two experiments are presented in Table 3. While the Generalization effect is significant for both experiments, the appropriate test for the predicted asymmetry of the mediation hypothesis is a *t* test, based on the within-*S*s error, of the differences between the Same Response and Different Response negative stimuli. This test was significant for Exp. II, ($P < .02$) and of borderline significance for Exp. I. ($P < .05$, one tail).

TABLE 3

ANALYSES OF VARIANCE OF ADJUSTED NUMBER OF CONDITIONED RESPONSE

Source	*df*	*F*	
		Exp. I	Exp. II
Between *S*s	23		
Tone arrangement (T)	1	.0005	.12
Hand arrangement (H)	1	.35	.88
T × H	1	4.84	.002
Error (b)	20	(37.85)[a]	(62.52)[a]
Within *S*s	48		
Generalization (G)	2	5.50*	48.03**
G × T	2	4.25*	.40
G × H	2	.97	1.93
G × H × T	2	.04	.46
Error (w)	40	(7.38)[a]	(8.66)[a]
Total	71		

[a] Numbers in parentheses are error mean squares.
* $P < .025$.
** $P < .001$.

The negative stimuli were identical in the two experiments. Table 4 is the summary of a combined analysis of the adjusted numbers of CRs to the two negative tones only. The significant effect for Experiments indicates the obvious finding of greater primary generalization from a tone to other tones than from a light to tones. Since responses to the CS were omitted from this analysis, the *F* for the Generalization effect is a direct test of the asymmetry of generalization. It was highly significant for the two experiments combined, and the *F* for the interaction of the Generalization effect with Experiments was less than unity. The significant Generalization × Tone Arrangement interaction was produced by the fact that there were

TABLE 4

ANALYSIS OF ADJUSTED NUMBER OF CRs TO NEGATIVE STIMULI IN EXP. I AND II

Source	*df*	*MS*	*F*
Between *S*s	47		
Tone arrangement (T)	1	1.55	.03
Hand arrangement (H)	1	63.99	1.35
Experiment (E)	1	485.19	10.21*
T × H	1	65.12	1.37
T × E	1	.93	.02
H × E	1	28.39	.60
T × H × E	1	49.22	1.04
Error (b)	40	47.51	
Within *S*s	48		
Generalization (G)	1	82.10	15.49**
G × T	1	53.02	10.00*
G × H	1	7.10	1.34
G × E	1	3.83	.72
G × T × H	1	5.19	.98
G × T × E	1	15.31	2.89
G × H × E	1	3.56	.67
G × T × H × D	1	3.00	.57
Error (w)	40	5.30	
Total	95		

* $P < .005$.
** $P < .001$.

somewhat more CRs to the high tone than to the low tone in the counterbalanced arrangement.

The assumption of the operation of both a facilitative effect and an associative, mediational effect leads to a prediction concerning the correlation across *S*s between mean reaction time and the number of CRs for the two negative stimuli. The facilitative effect should tend to produce a negative correlation between these variables for both tones. For the Same Response stimulus, where the mediated equivalence effect is supposed to be operating, there should be a second source of negative correlation. That is, the earlier the mediating response occurs, the greater its chance of eliciting the CR within the scoring interval. This should not be the case, however, for the Different Response stimulus. This source of negative correlation should be absent, or, if a mediated distinctiveness effect is operating, there should be a source of positive correlation. This reasoning leads to the rather clear prediction that there should be a higher negative correlation between reaction time and the number of CRs for the Same Response stimulus. For Exp. I this correlation was −.24 for the Same Response tone, and +.18 for the Different Response tone. The corresponding correlations for Exp. II were −.62 and −.44, respectively. Both differences were in the predicted direction. Employing the test for significance of a difference between related correlation coefficients, neither difference was quite significant. However, when the probabilities were combined by the Fisher-Pearson chi square procedure, the difference was significant for the two experiments combined ($P = .044$, two-tailed test).

DISCUSSION

It is the opinion of the authors that the results of these experiments should be interpreted with considerable caution. Since the facilitative effect is obviously a powerful one while the mediational one is relatively small, the danger of artifactual results must not be ignored. But in view of the important role assigned to mediated generalization in the theoretical literature, it seems important to ascertain whether it does operate at the level of the conditioned response, which is commonly regarded as a rather fundamental form of associative learning. The analyses of the conditioning measures, adjusted for their correlation with reaction time, appear to provide moderately good evidence that a mediational effect was operating in these experiments as well as the facilitative one. The verified prediction concerning the higher negative correlation between generalization and reaction time for the "equivalent" stimulus as opposed to the "distinctive" one provides corroborative evidence. The acceptance of the null hypothesis concerning the interaction of the generalization effect with experiments is in accord with the prediction that the amount of mediated generalization should be independent of whether or not the generalized stimuli affect the same sensory modality as the CS.

It is obvious to suggest that mediation should be studied in the absence of facilitative effects, since statistical control of extraneous variables is never completely satisfactory. In some instances this may be possible. Yet if one wishes to study the associative consequences of a response at the time of its occurrence, one must be prepared to consider its other contemporary effects. This is not a problem specific to the present eyeblink experiments.

The strong facilitative effect which has been found in this series of experiments is of some interest and importance in its own right. Of immediate interest, is the role which it has probably played in other eyeblink studies which have involved concurrent motor responses during conditioning. For example, the study of Voeks (1954), purporting to show step-wise conditioning curves, involved a sort of two-hand reaction-time response which she conceived to be part of the true CS. The closure of two keys presented an auditory CS, the keys then being held down against a "stiff spring" for the duration of this signal. Responses were counted as CRs if they occurred anytime after the onset of the auditory stimulus. It would be difficult, indeed, to design a situation better calculated to produce a facilitative effect. This is, in part, evidenced by the fact that her "conditioning" level reached approximately 35% on Trials 2 and 3 and 60% on Trials 4 and 5. It is extremely doubtful that this situation involved conditioning to any important degree. Similar complications would be involved in such studies as those of Miller and Cole (1936) and Beck (1939).

SUMMARY

In two differential eyeblink conditioning experiments, *S*s made more CRs to negative stimuli accompanied by the same manual response as the CS, than they did to stimuli to which they made a different manual response. The faster the reaction time of the manual response, the more probable a CR, irrespective of which stimulus was involved. This was interpreted as evidence for a facilitation effect. The asymmetry of generalization was still present, however, when the conditioning measures were adjusted for correlation with reaction time. This was interpreted as indicating the probable action of an associative, mediated generalization effect. The correlation between number of CRs and reaction time was significantly more negative for the "equivalent" generalized stimuli than for the "distinctive" stimuli. This was also interpreted as indicating the presence of both facilitative and mediational effects.

There was no evidence that the magnitude of the mediation effect was different when generalization was from one tone to another than when it was from a light to a tone.

Implications of the facilitative effect for other eyeblink studies involving concurrent motor responses were indicated.

REFERENCES

BECK, L. H. Conditioning and the coordination of movements. *J. gen. Psychol.*, 1939, **20**, 375–397.

GRICE, G. R., & DAVIS, J. D. Mediated stimulus equivalence and distinctiveness in human conditioning. *J. exp. Psychol.*, 1958, **55**, 565–571.

MILLER, J., & COLE, L. E. The influence of a "voluntary" reaction upon the development and the extinction of the conditioned eyelid reaction. *J. genet. Psychol.*, 1936, **48**, 405–440.

VOEKS, V. W. Acquisition of S-R connections: A test of Hull's and Guthrie's theories. *J. exp. Psychol.*, 1954, **47**, 137–147.

(Received June 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

ROLES OF ASSOCIATION VALUE AND SYLLABLE FAMILIARIZATION IN VERBAL DISCRIMINATION LEARNING

WILLARD N. RUNQUIST[1] AND MADELYN FREEMAN

Hobart and William Smith Colleges

Paired-associate and serial verbal learning may be conceived of as consisting of two stages. The first stage consists of delimiting the population of responses, i.e., learning what the responses are. The second stage consists of pairing each response with its correct stimulus or learning the correct order of these responses. Although logically the first stage must precede the second, it is likely that the two occur concomitantly in the ordinary learning situation.

Several studies of the role of association value and familiarization of syllables in paired-associate and serial learning have suggested that factors associated with responding are more important than "perceptual" factors in acquisition of verbal lists (e.g., Cieutat, Stockwell, & Noble, 1958; Mandler & Campbell, 1957; Noble, 1952, 1955; Noble & McNeely, 1957). One hypothesis which may be advanced concerning these results is that the primary influence of association value and familiarization is in the first stage (the response-learning phase) of acquisition. According to this hypothesis, the letters making up a nonsense syllable must be integrated into a single response before correct anticipations can occur (Mandler, 1954). Syllables of high association value and familiarity are conceived of as representing verbal habits established outside the experiment and thus are more readily integrated. The number of responses available on any given trial is then determined by how far this integration stage has progressed.

However, suppose a learning task is used in which the total number of available responses is determined by the task, rather than the rate of response integration. This may be accomplished by using a procedure called verbal discrimination learning, where *S* is presented with a pair of syllables and must select the response arbitrarily designated as "correct" from these two. The response population from which *S* must make his choice is thus restricted to the two syllables presented, and integration in the sense described above is unnecessary. It might then be expected that neither association value nor familiarization would have much effect upon acquisition in this task.

The following experiments are designed to investigate the role of both association value and familiarization on the verbal discrimination learning of nonsense syllables.

EXPERIMENT I

The first two experiments were concerned with the effects of association value. Two methods of varying association value are available for verbal discrimination learning: A pair may be made up of one high and one low association value syllable, or the two syllables making up a pair may both be high or both low association value. According to the hypothesis, there seemed to be no reason to favor one method or the other. Although this conclusion later appeared to be in error, for the first experiment the first method (heterogeneous pairs) was used.

---

[1] Now at Pomona College.

Method

Design and materials.—Two sets of eight pairs of syllables were constructed from Glaze's lists (1928). In each pair one syllable was 80% to 100% association value and one syllable was 0% to 27% association value. Formal similarity was kept to a minimum both within and between sets, and the same number of letters was duplicated within a set. These sets were then combined to make a list of 16 pairs of syllables. For half of the *S*s, the high association value syllables of Set I and the low association syllables of Set II were correct. For the other half of the *S*s, the low association value syllables of Set I and the high association value syllables of Set II were correct. This reduced the chances of extraneous factors such as differential difficulty or interference with specific syllables influencing the results. It was felt that having both high and low association value syllables correct in the same list would avoid *S*s recognizing, for example, that all the "word-like" ones were correct.

Procedure.—The 16 pairs in each list were arranged in three orders and presented with a 16-mm. filmstrip projector. The pairs were flashed from the back of the screen so that *S* could not see *E* or the projector during the presentation. General illumination was reduced to minimized distraction. Syllables appeared black on a white background and were ⅜ in. high. They appeared on an 8 × 8-in. milk glass piece in the center of a 3 × 3-ft. wooden screen. The two syllables were presented one above the other for 2 sec. followed by the correct syllable for 2 sec. The intertrial interval was 4 sec. The position of the syllables in a pair varied from trial to trial. All *S*s were given 15 trials unless they went three consecutive trials with no error. They were then stopped and given a perfect score for the remaining trials. The *S*s, 20 volunteers from Hobart and William Smith Colleges, were instructed to select and spell aloud the syllable designated as correct on each presentation including the first trial. All *S*s were naive to verbal learning experiments.

Results

The percentages of correct anticipations for the high association value syllables and low association value syllables on each trial are shown in Fig. 1. The hypothesis that variations in association value would not produce performance differences is not supported by these data. Statistically, the mean total correct anticipations for the high association value syllables is significantly greater than for the low association value syllables ($t = 3.65$). Seventeen of the 20 *S*s gave more correct responses to the high than to the low association syllables.

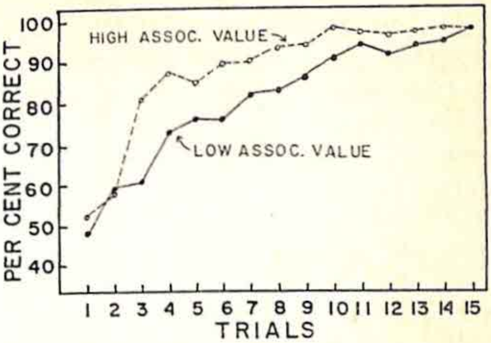


FIG. 1. Percentage of correct responses on each trial for high and low association value syllables in Exp. I (heterogeneous pairs).

EXPERIMENT II

Rather than completely reject the hypothesis on the basis of this single experiment, explanation was sought in terms of other factors associated with making the response. The facilitation in performance obtained here may have been due to greater ease of responding with the high association syllables. Essentially this notion is what Noble (1955) has called "motor patterning." If it is easier to learn to make the high association value responses than the low association value responses, some facilitation might be expected. This is a kind of response integration although only the motor aspect is being referred to here. If the pairs, however, are constructed of two high or two low association value syllables, greater ease of responding to high association value syllables cannot determine which response is made, since one of the two responses must be made each time a pair is presented. The second study, then, compared verbal discrimination learning in which pairs were homogeneous, i.e., either both syllables had high or both syllables had low association value.

Method

The same syllables were used as in Exp. I, except that they were re-paired to form two sets of eight syllables in which Set I contained only high association value syllables and Set II only low association value syllables. Duplicated letters were again kept at a minimum and equalized within each set. Two lists were used consisting of the two sets combined, with the correct member of each pair in List 1 being the incorrect member of the same pair in List 2. Half of the *S*s learned each list.

The *S*s were 20 volunteers from Hobart and William Smith Colleges. All were naive with respect to verbal learning. The procedure was identical to that in Exp. I.

Results

The acquisition curves for the high and low association value pairs are presented in Fig. 2. Although there is some evidence for better performance on the high association value pairs, the differences are much smaller than those obtained in Exp. I. Comparison of the mean total correct responses for the two kinds of syllables resulted in a *t* of 2.07, which is just short of significance at the .05 level for 19 *df*. Although a constant number of trials was used, all *S*s obtained all eight items within each set correct on at least one trial. Thus trials to 8/8 criterion may be used as a measure of acquisition rate. Seventeen of the 20 *S*s reached this criterion first on the high association value pairs. A comparison of the mean trials to this criterion for the two kinds of pairs resulted in a *t* of 3.45. It thus appears that the difference in performance shown in Fig. 2 is reliable, although a small one.

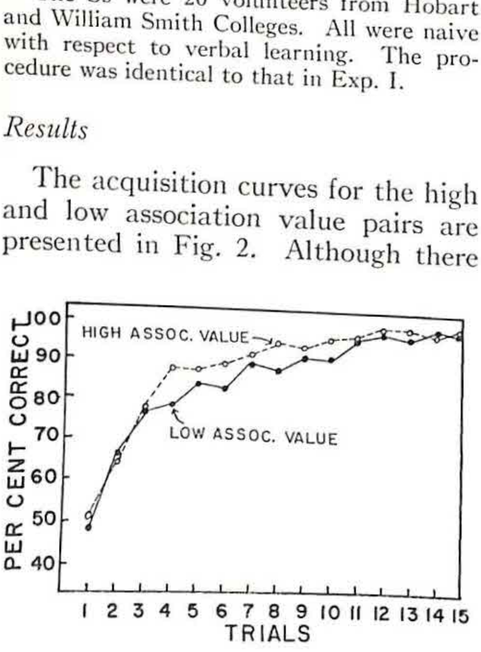


FIG. 2. Percentage of correct responses on each trial for high and low association value syllables in Exp. II (homogeneous pairs).

A comparison was made between Exp. I and II of the mean difference between the total correct responses to high and low association value pairs. The value of *t* obtained was 1.73 which is only significant at about the .08 level.

The results of Exp. I and II do not support the hypothesis that the facilitative effects of association value can be attributable solely to the response-learning phase. Although there are several interpretations of these differences, the one favored here is suggested by Underwood and Richardson (1958). High association value syllables may provide some sort of extra cue value, often called "meaning," thus making them less susceptible to interference from duplicated letters. Since in the lists used in these experiments it was impossible to avoid some duplication of letters, it is not unreasonable to assume that the performance on the low association items in Exp. II was depressed by the slight amount of interference present. Some support for this position comes from the fact that a pilot study using 10 homogeneous pairs in which there were no duplicated letters, except a few vowels, produced no differences between high and low association value pairs ($t = 1.27$ with 19 *df*). The present experiments used lists of 16 pairs because it was felt that acquisition was too rapid with a 10-item list to give any differences a chance to appear. Increasing the length of the list necessarily increases the duplicated letters due to the restricted number of letters and syllables available, thus increasing interference. The difference obtained in Exp. II, while not significantly smaller than that obtained in Exp. I, tends to support the hypothesis that the difference in the latter experiement was mostly due to differential ease of making the response.

EXPERIMENTS III AND IV

The purpose of these two studies was to compare the effects produced by the familiarization of nonsense syllables (Noble, 1955) with those produced by association value.

Method

Design and materials.—The rules for the construction of lists were the same as those in the previous experiments, with the exception that all syllables were from 0% to 27% association value (Glaze, 1928). The number of duplicated letters within a set was slightly higher for these lists than in Exp. I and II. As in the previous studies, a given syllable in each pair was correct for half the *S*s and incorrect for the other half.

Procedure.—Prior to learning the list each *S* was given familiarization training on half the syllables.

Half the syllables were correct in learning and half the syllables were incorrect For Exp. III, one syllable was chosen from each pair; for Exp. IV, both syllables of half of the pairs were used. Two different familiarization lists were employed in each experiment so that syllable difficulty, etc., was partially counterbalanced between the familiarized and nonfamiliarized syllables. In familiarization training each syllable was exposed singly on the screen for 2 sec. The *S* was instructed to spell out each syllable as it appeared. The 16 syllables were arranged in six different orders, but there was no interval between orders. No syllable preceded or followed any other syllable more than once. Each syllable was seen and spelled 12 times by each *S*.

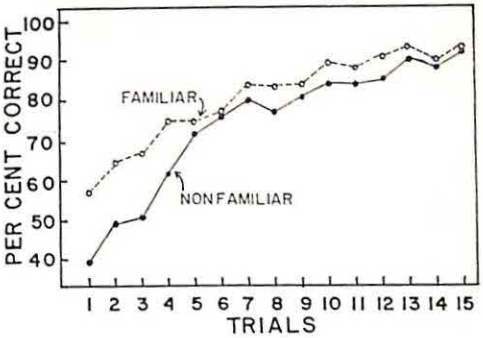


FIG. 3. Percentage of correct responses on each trial for familiar and nonfamiliar syllables in Exp. III (heterogeneous pairs).

The verbal discrimination learning procedure was identical with that in Exp. I and II. Twenty volunteers from the Hobart and William Smith students served as *S*s in each study. None of the *S*s had served in a verbal learning experiment before.

Results

The percentage of correct responses on each trial for familiarized and nonfamiliarized syllables in Exp. III is shown in Fig. 3. With heterogeneous pairs performance level is higher when the familiarized syllable is correct than when it is incorrect,

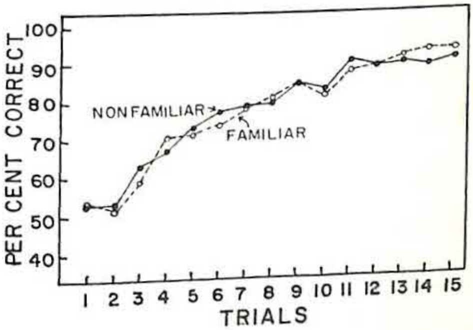


FIG. 4 Percentage of correct responses on each trial for familiar and nonfamiliar syllables in Exp. IV (homogeneous pairs).

with the advantage being particularly marked in the early trials. A *t* test comparing the mean total correct responses for the two kinds of syllables resulted in a value of 3.78 ($P < .01$).

The acquisition curves for Exp. IV are presented in Fig. 4. In this experiment, however, there is no difference in performance between pairs of familiarized and nonfamiliarized syllables. Comparison of the mean total correct responses gave a *t* of .33, which is clearly not significant.

DISCUSSION

The effects of familiarization training on verbal discrimination learning are not the same as the effects of association value. The fact that familiarization produced facilitation in the case of heterogeneous pairs is consistent with the effects produced by association value, although there are some slight differences. In the case of association value, performance on the two kinds of pairs starts at the same level, then diverges before converging again near the asymptote. Familiarization, on the other hand, produces large differences on the first trial. This large initial difference in the case of familiarization is probably produced by a tendency to guess and to keep guessing the syllables which have just been seen. This is clearly shown by the fact that performance on the nonfamiliarized syllables begins below the chance level (at 39%), while performance on the familiarized syllables begins at 57%. These results are also consistent with the hypothesis attributing facilitation to differential ease of responding for the two kinds of syllables. Since the differential ease of responding might not be established until after a few trials when association value was varied, facilitation did not take place until later in learning, whereas familiarization training established these habits prior to learning the list.

The striking result of these studies, however, is the failure of familiarization training to produce differences in performance when homogeneous pairs are used. This result is also consistent with the interpretation given the association value experiments. If the differences which occurred in the homogeneous pairs association value experiment are attributed to the cue value of association or "meaning," familiarization training should not produce differences since it can be shown that familiarizing operations, in general, do not increase association value (Riley & Phillips, 1957). It might be argued that a difference would have been obtained in Exp. IV had familiarization training been continued. This cannot, of course, be denied; however, the 12 trials given here were sufficient to produce large differences in Exp. III when conditions for producing these differences were allowed. In fact, the purpose of Exp. III was to ensure that the familiarizing operations were sufficient to produce performance differences in one situation where these differences might be expected.

It is felt that, over-all, the results support a response-oriented interpretation of the effects of association value and familiarization, although they do indicate some cue effect of association value. The results of Exp. IV represent the only real support for the original predictions from the response integration hypothesis that association value and familiarization would not facilitate verbal discrimination learning. Nevertheless, an attempt has been made to show that complicating factors were present in the case of heterogeneous pairs.

SUMMARY

Four verbal discrimination learning experiments were performed to test the hypotheses that association value and familiarization with the syllables would not facilitate performance. Two experiments varied association value and two gave familiarization training on the syllables. In one experiment each pair of items to be learned consisted of one high and one low association value syllable (heterogeneous pairs), while in the other experiment, both syllables were either high

or low association value. The same held true for the familiarization studies.

The results showed that both association value and familiarization facilitated performance with the heterogeneous pairs, while only association value facilitated performance with the homogeneous pairs. These results were discussed in relation to an interpretation of the role of association value and familiarization in learning in terms of their effects on response availability

REFERENCES

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m') in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

MANDLER, G. H., & CAMPBELL, E. H. Effects of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 269–273.

NOBLE, C. E. The role of stimulus meaning (m) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446.

NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

NOBLE, C. E., & MCNEELY, D. A. The role of meaningfulness (m) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

RILEY, D. A., & PHILLIPS, L. W. The influence of syllable familiarization on association value, serial rote learning, and reminiscence. *Amer. Psychol.*, 1957, **12**, 425. (Abstract)

UNDERWOOD, B. J., & RICHARDSON, J. Studies of distributed practice: XVIII. The influence of meaningfulness and intralist similarity of serial nonsense lists. *J. exp. Psychol.*, 1958, **56**, 213–219.

(Received June 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

EFFECT OF RUNWAY SIZE AND DRIVE STRENGTH ON ACQUISITION AND EXTINCTION[1]

DONALD J. LEWIS
Louisiana State University
AND JOHN W. COTTON
Northwestern University

The theoretical importance of competing responses has been repeatedly emphasized in the past (Guthrie, 1935; Wendt, 1936), and data are beginning to accumulate indicating that competing responses do affect ordinarily measured behaviors in important ways. Cotton (1953) showed that a major part of the effect of drive on asymptotic running speed was due to the intervention of competing responses. When trials on which competing responses occurred were not analyzed, drive did not have as great an effect on performance as it did when competing response trials were included in the analysis. Pereboom and Crawford (1958) showed essentially the same effect for magnitude of reward and number of acquisition trials.

Adelman and Maatsch (1955, 1956) have shown that controlling the competing responses during extinction results in a marked effect upon behavior. When they allowed *S*s to climb out of the goalbox, running speeds were increased during extinction. This finding was confirmed by Cotton, Lewis, and Metzger (1958) and extended to acquisition behavior for both the runway and the **T** maze.

Sheffield (1950), Stanley (1952), Amsel (1958), and others have emphasized the drive-arousing properties of the extinction operation, and others (Spence, 1956; Taylor, 1951) have indicated the connection between drive and competing responses.

There seems little doubt, then, about the importance of competing responses. And the effect of competing responses in conjunction with drive is also of interest. The present paper is a report of an experiment in which two conditions of drive and two conditions of competing responses were manipulated factorially. The drive conditions were established by hours of food deprivation; the competing response conditions were established by two runway sizes, large (L) and small (S). It was assumed that the L condition would permit many competing responses, whereas the S condition would tend to prevent the competing responses from occurring. In the L runway *S* could turn around, scratch himself, move from one side of the alley to the other, and perform many responses other than running to the goalbox. The S runway, it was believed, would prevent most of these competing responses.

METHOD

Subjects.—The *S*s were 80 naive male albino rats of the Sprague-Dawley strain. They varied from 66 to 81 days of age, with a mean of 74.3 days. Because it was necessary to make structural adjustments in the apparatus for different experimental groups, departures from purely random assignment of *S*s were necessary. When *S*s arrived, they were assigned by a table of random numbers to their individual living cages. Then they were assigned to the experimental conditions in blocks of four with each block of four serving under identical conditions. Four *S*s were run in each of the four experimental conditions before any one of the experimental conditions received its second assignment of four *S*s. Replications continued in this fashion until 20 *S*s were assigned to each of the experimental conditions. Four *S*s were discarded because of a procedural error, and they were replaced.

---

[1] This study was supported by a grant from the National Science Foundation.

The writers wish to express their appreciation for the work of Glen Jensen who participated in all phases of the present investigation.

Apparatus.—The apparatus was a straight runway, the inside dimensions of which could be altered by putting in or taking out Plexiglas inserts. Two adjustments of the inside dimensions were employed. In one case the runway was set at its maximum width and depth. In this L condition the inside dimensions of the runway proper were 8½ in. high, 4 in. wide, and 4 ft. long. The inside dimensions of the goalbox were 8½ in., 9 in., and 10½ in. The goalbox also contained a recess, 2¾ in. by 1¾ in., in which the foodcup was placed so that it could not be seen before a complete goalbox entry. For the startbox the dimensions were 8½ in., 4 in., and 6 in. The inside dimensions of the startbox for the S condition were 2½ in., 1⅗ in., and 6 in. For the runway proper the dimensions were 2½ in., 1⅗ in., and 4 ft. The goalbox was 2½ in. high but remained unchanged in the other dimensions. The interior of the apparatus was painted flat black. Response time was defined as the interval required from the opening of the startbox door until *S* broke a photoelectric beam 2 in. from the goalbox. The distance between the startbox and the photoelectric beam was 44½ in. Times were taken automatically by a Standard Electric timer, calibrated in .01 sec.

Design.—Two variables were manipulated factorially. One was the number of hours of food deprivation and the other was the inside dimensions of the apparatus. One-half of the *S*s were assigned to the L runway for acquisition and extinction, and one-half were assigned to the S runway. One-half of the *S*s run in the L apparatus were deprived of food for 2–3 hr. at the beginning of any experimental session. The remaining half of the *S*s were deprived approximately 22 to 23 hr. The same deprivation conditions were applied to *S*s run in the S apparatus.

Deprivation and taming.—All *S*s were placed on their appropriate deprivation schedule at least seven days before the initiation of experimentation. The *S*s were handled for at least 5 min. on each of these seven days and made familiar with the 45-mg. lab chow pellets to be used during the acquisition phase of the experiment.

Acquisition.—Each *S* received 10 trials a day for five days. Then two days intervened during which *S* received no trials, but was maintained on the appropriate deprivation schedule. There followed another 10 trials a day for five days. A total of 100 acquisition trials was given with approximately a 15-sec. intertrial interval and with a 45-mg. pellet available in the goalbox on each trial. The *S*s were detained in the goalbox approximately 10 sec. on each trial.

Extinction.—Extinction began two days after the one-hundredth acquisition trial. During these two days *S*s were maintained on their appropriate deprivation schedule. Ten nonreinforced trials a day for five successive days were given during the extinction period. If on one of these trials *S* did not reach the goalbox within 1 min., he was removed from the apparatus and the response was recorded as 1 min. The intertrial interval remained at approximately 15 sec., and *S*s continued to be detained in the goalbox for approximately 10 sec. on each trial.

RESULTS

Acquisition.—The reciprocal of the running time (speed) for each trial was determined for each *S*, and all analyses were performed on these speeds. A graph of the mean speed by 10-trial blocks is presented in Fig. 1. An over-all analysis of variance was performed on the means

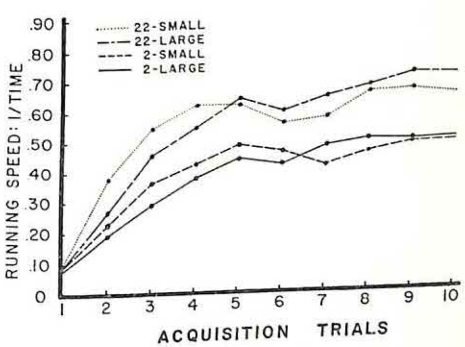


FIG. 1. The relationship between acquisition trials and running speed. Each of the 10 points on the abscissa represents the mean of 10 trials. Speeds are the reciprocals of running time in seconds.

of the 100 acquisition trials. The apparatus size was found to be *not* an effective variable ($F = .32$, $df = 1/76$). Drive was significant ($F = 15.18$, $df = 1/76$); and there was no interaction between apparatus size and drive ($F = .03$, $df = 1/76$). Figure 1 shows the clear superiority of the high-drive groups throughout acquisition. It also seems clear from Fig. 1 that the two drive conditions resulted in different asymptotes. The slight dip in all four acquisition curves just after Trial 50 can probably be explained by the two-day interval after Trial 50 during which *S*s received no training.

The over-all analysis of variance, which shows apparatus size not to be an effective variable, is probably not completely appropriate for these data because the analysis includes values for all trials in the computations of means for each group. The initial elevation of the S groups is counterweighted by the later elevation of the L groups, and the over-all analysis is insensitive to this change as a function of trials. Therefore, another analysis of variance was performed with trials as an added variable. The means of 10-trial blocks 31–40, 41–50, 51–60, and 61–70 were taken for each *S* and these scores made up the added variable of trials. This analysis confirms what Fig. 1 seems to show: that the effect of apparatus size is a function of the number of acquisition trials. The *F* for the interaction between apparatus size and trials was 5.53. A value of 3.97 is required for significance at the .01 level with 3 and 114 *df* (the pooled *S*s × Trials served as an error term).

Perhaps, because of its *post hoc* nature, this "repeated measurements" analysis capitalized on observed chance variation in the data. The alternative is to be blinded by a poor first guess as to the appropriate analyses. The final test of any empirical finding is its reliability over several experiments.

A record was kept of all the trials on which competing responses occurred. A competing response was defined as any behavior which did not result in forward progress through

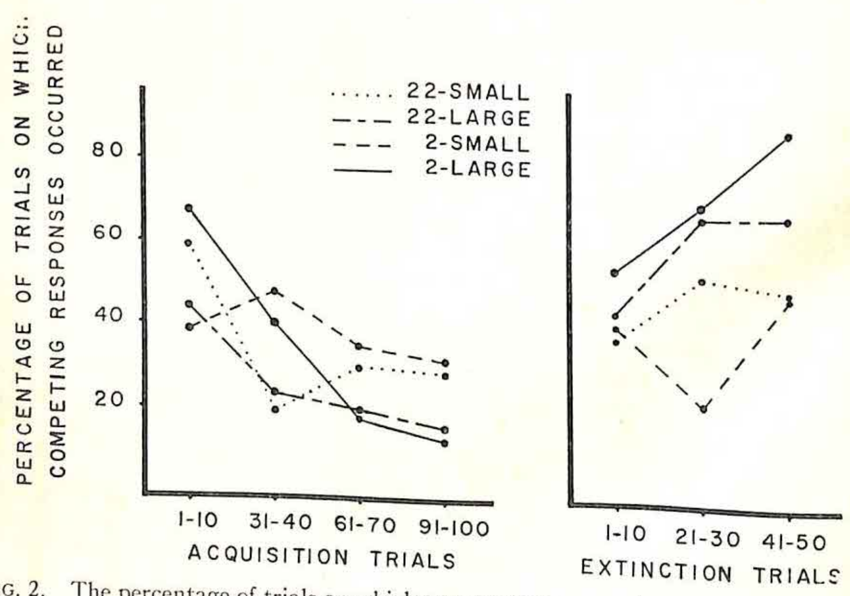


FIG. 2. The percentage of trials on which one or more competing responses occurred.

TABLE 1

ANALYSIS OF THE CHANGE IN COMPETING RESPONSES WITH TRIALS DURING ACQUISITION AND EXTINCTION

Source	*df*	Acquisition		Extinction	
		MS	*F*	*MS*	*F*
Drive (D)	1	46.44	2.07	18	—
Apparatus size (A)	1	15	—	217	11.88*
D × A	1	7	—	7	—
Between *Ss* in the same group	76	22.25		18.4	
Trials (T)	3	141.6	33.2*	108.5	30.74*
T × D	3	21.3	4.99*	1.56	—
T × A	3	18.3	4.28*	27	7.65*
T × D × A	3	36.3	8.50*	4	1.13
Pooled *Ss* × T	228	4.3		3.53	

* $P = .01$.

the maze. Such behaviors as retracing, face washing, and scratching were ruled as competing responses. Figure 3 shows the percentage of trials on which one or more competing responses occurred as a function of the manipulated variables, and Table 1 shows the result of "repeated measures" analysis of these data.

During acquisition none of the variables resulted in significant differences by themselves, but they both interacted significantly with trials. In addition, there was a significant triple interaction of the two variables with trials. If the mean square for the triple interaction is used as the error terms in testing the other differences, only the F for Trials remains significant. The Trials F means that the number of competing responses decreases as acquisition proceeds.

The triple interaction seems to be attributable to the differential effect over trials of low drive, depending on the apparatus that is used. In the L apparatus the low-drive group yields a higher percentage of trials with competing responses early in acquisition than does the high-drive group. Late in acquisition, however, the low-drive group in the large apparatus has very few competing responses, whereas the low-drive group in the small apparatus has many.

To state the meaning of this triple interaction in another way: Apparatus has no differential effect over trials for the high-drive group. And, in general, fewer competing responses occur in the S apparatus throughout acquisition. The effect of drive, however, is reversed from early to late acquisition for the L apparatus groups. Early in acquisition more competing responses occur in the L apparatus, and late in acquisition more competing responses occur in the S apparatus. Decreasing apparatus size decreased the percentage of trials on which competing responses occurred only very early in acquisition for the low-drive group. Other than that, more competing responses occurred in the S apparatus.

Extinction.—An over-all analysis of variance of the extinction running speeds was performed. Both Drive ($F = 19.1$ $df = 1/76$) and the Apparatus ($F = 13.5$, $df = 1/76$) were highly significant variables. There was no interaction between the two ($F = .05$, $df = 1/76$). The interpretation of this analysis can be obtained from Fig. 3. This figure indicates that the S apparatus resulted in *increased* resistance to extinction, and that the greater drive also resulted in prolonged extinction. An analysis of covariance in which the mean of the last 10 acquisition trials was used as the control variable and the mean of all extinction trials as the prediction variable gave approximately the same results. The F for the apparatus was increased to 28.86 for 1 and 75 *df*, however. An analysis of covariance, again with the mean of the last 10 acquisition trials as the control variable, but with the mean of the last 10 extinc-

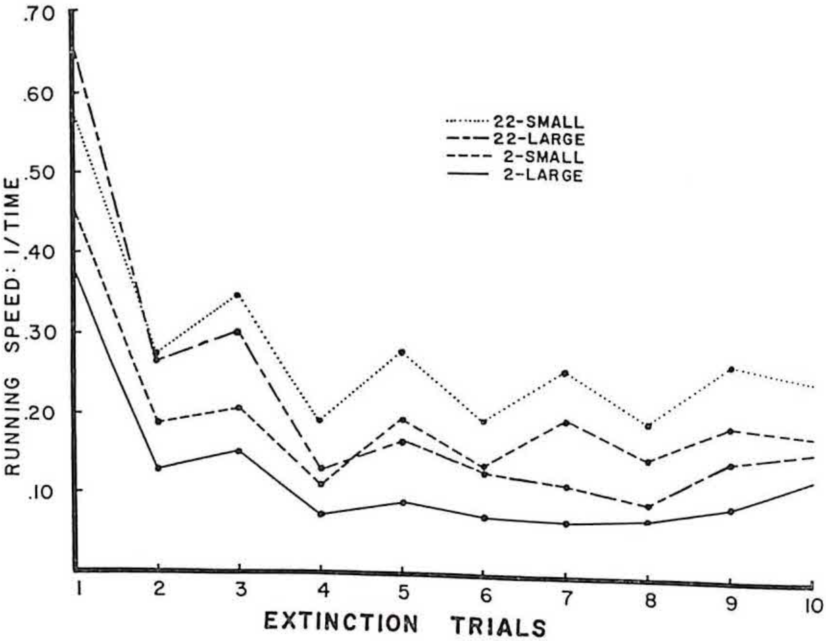


FIG. 3. The relationship between extinction trials and running speed. Points 1 to 2, 3 to 4, 5 to 6, 7 to 8, and 9 to 10 represent blocks of 10 extinction trials given on each of five days. Points 2 to 3, 4 to 5, 6 to 7, and 8 to 9, represent spontaneous recovery between 10-trial extinction blocks.

tion trials as the predicted variable, shows the Apparatus as remaining significant ($F = 10.10$, $df = 1/75$).

Finally, a repeated measurements analysis was performed using the mean of the last 10 acquisition trials, the mean of extinction Trials 21 to 25, which form the midpoint of the extinction trials, and the mean of the last 5 extinction trials as points for the trend. A summary of this analysis is presented in Table 2. Drive alone remains significant; Apparatus is not significant. Because all the other extinction analyses show Apparatus to be a highly significant variable, this low F value is perhaps surprising. The reasons for the nonsignificant F for Apparatus can be seen by looking at Fig. 1 and 3 jointly. At the end of acquisition, shown in Fig. 1, the mean of the S conditions is below that of the L conditions. During extinction, shown in Fig. 3, just the reverse is true; the mean of the S conditions is above that of the L conditions. This reversal cancels out the over-all effect of the apparatus. The reversal also means that there should be a significant interaction of apparatus with trials. The very significant F

TABLE 2

ANALYSIS OF THE CHANGE IN EXTINCTION RUNNING SPEEDS WITH TRIALS

Source	df	MS	F
Drive (D)	1	.741	17.50*
	1	.157	3.68
Apparatus size (A)	1	.001	.03
D × A	76	.043	
Between Ss in the same group			
Trials (T)	2	4.556	236.32*
T × D	2	.100	5.20*
T × A	2	.122	6.37*
T × D × A	2	.005	.28
Pooled Ss × T	152	.019	

* $P = .01$.

for the Trials × Apparatus interaction confirms this interpretation. The S apparatus resulted in slower running speeds at the end of acquisition but in faster running speeds throughout most of extinction.

The interaction between drive and trials is also significant. This significance is attributable to the general decrease in the effectiveness of drive as extinction progresses.

Figure 3 and an analysis of variance show that during extinction more competing responses occurred in the L apparatus, and that drive had no effect on competing responses.

Spontaneous recovery.—Ten extinction trials were run on the first extinction day, and 10 trials a day were given for four succeeding days. To determine the amount of spontaneous recovery, the *difference* between the speeds on the tenth extinction trial and the eleventh extinction trial was computed, and an analysis of variance performed on these difference scores. Drive was very significantly effective ($F = 10.4$, $df = 1/76$), Apparatus size was not ($F = 1.2$), and there was no significant interaction between the two ($F = .14$). Figure 3 shows clearly that the higher the drive, the more the spontaneous recovery.

DISCUSSION

As in a study by Cotton, Lewis, and Metzger (1958), a wide runway was accompanied by an increased number of trials on which competing responses occurred. Because more competing responses usually meant slower running times, rapid extinction corresponded with an increase in competing responses. The two measures are, however, to some extent independent. The *S*, for example, can run fast or slow on trials on which no competing responses occur. Also, it is possible for a competing response to occur on a fast trial as well as on a slow trial. Therefore, it is meaningful to conclude that competing responses are at least one determiner of extinction, as measured by decreased running times, and this conclusion is in line with a number of interference theories of extinction.

That the S apparatus, in general, increased the number of competing responses was a surprise, although there was at least a suggestion of such a result in the previous Cotton, Lewis, and Metzger study. Evidently apparatus size has a different effect on competing responses during acquisition than during extinction. And it is clear that a small apparatus, little bigger than a rat's body, does not necessarily curtail competing responses. The initial plan to manipulate competing responses by restricting the area for locomotion was a failure.

SUMMARY

Two conditions of drive—2 and 22 hr. of food deprivation—were combined factorially with a large and a small runway in an attempt to determine some of the effects of drive on competing responses. The length of the runway was the same in both apparatus conditions, but the small runway was constricted greatly in other dimensions in an attempt to reduce competing responses. Twenty animals were assigned to each of the four conditions. They were given 100 acquisition and 50 extinction trials. The *S*s were run under the same experimental conditions during extinction that they had experienced during acquisition.

The major findings were:

1. High-drive resulted in faster running speeds throughout acquisition and early extinction.
2. Early in acquisition the small apparatus resulted in faster speeds; there is a reversal of this trend later in acquisition.
3. The small apparatus decreased the number of trials on which competing responses occurred only for the first few trials of acquisition and then only for the low-drive groups. Other than this, more competing responses occurred in the small apparatus during acquisition.
4. During most of extinction the small apparatus produced greater running speeds than the large, although this superiority was lost by the end of extinction.

5. The high-drive condition resulted in greater spontaneous recovery.

6. In extinction, the small apparatus served to reduce competing responses.

REFERENCES

ADELMAN, H. M., & MAATSCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, 1955, **50**, 61–65.

ADELMAN, H. M., & MAATSCH, J. L. Learning and extinction based upon frustration, food reward, and exploratory tendency. *J. exp. Psychol.*, 1956, **52**, 311–315.

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

COTTON, J. W., LEWIS, D. J., & METZGER, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935

PEREBOOM, A. C., & CRAWFORD, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56**, 82–85.

SHEFFIELD, V. F. Resistance to extinction as a function of distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 303–313.

SPENCE, K. W. Behavior theory and conditioning. New Haven: Yale Univer. Press, 1956.

STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.

WENDT, G. R. An interpretation of inhibition of conditional reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258–81.

(Received June 26, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

LEARNING AND RETENTION OF PAIRED ADJECTIVES AS A FUNCTION OF PERCENTAGE OCCURRENCE OF RESPONSE MEMBERS

RUDOLPH W. SCHULZ[1] AND WILLARD N. RUNQUIST[2]

Northwestern University

If the paired-associate verbal learning situation is considered an analogue of the classical conditioning situation, then the presentation of the response member of the S-R pairs corresponds to the presentation of the UCS and reinforcement. Therefore variation in percentage occurrence of response members in the paired-associate situation should be analogous to variation in the percentage of reinforcement in the conditioning situation.

At the time the present study was begun, percentage occurrence of response members (hereafter abbreviated ORM) had not been investigated in paired-associate verbal learning. Recently, Goss, Morgan, and Golin (1959) have reported that the mean number of trials to learn paired nonsense-syllable lists to criterion decreased as percentage ORM increased. Adjective lists were used in the present experiment to investigate the ORM dimension at five points; namely, 20%, 40%, 60%, 80%, and 100% ORM.

It is well established that intermittently reinforced responses are typically more resistant to extinction than continuously reinforced responses. If an unlearning (Melton & Irwin, 1940; Melton & von Lackum, 1941) extinction-like (Underwood, 1948) process underlies forgetting, at least in part, then it is conceivable that a paired-associate list learned with intermittent ORM will be more resistant to forgetting than a list learned with continuous ORM. Therefore a second purpose of the present experiment was to study retention as a function of percentage ORM.

METHOD

List construction.—All *S*s in all conditions learned the same 10 pairs of adjectives. Intralist similarity among these pairs was low. Details concerning the original selection of the pairs may be found elsewhere (Underwood, 1951). Five lists, differing only with respect to percentage ORM, were constructed utilizing these 10 pairs of adjectives. The method of manipulating percentage ORM was essentially identical with that used by Goss, Morgan, and Golin (1959); hence there is no need to detail it here. In short, the occurrence of the response members of the adjective pairs was scheduled so that percentage occurrence per trial for a given list was the same as the percentage occurrence for a given pair in the list within successive blocks of five trials. For example, with 20% ORM, 2 of the 10 stimuli were accompanied by their response members on each trial. The remaining eight stimuli were accompanied by a blank. On one of the five trials within each successive block of five trials each stimulus was accompanied by its response member once or 20% of the time. While Goss et al., used an 8-item list of nonsense syllables with 25%, 50%, 75%, and 100% ORM, the present study employed a 10-item list of adjectives with 20%, 40%, 60%, 80%, and 100% ORM.

Subjects and procedure.—The 20%, 40%, 60%, 80%, and 100% ORM lists were learned by five different groups of 50 *S*s each. The 250 *S*s were students at Northwestern University taking introductory psychology courses. They were naive to the learning of paired adjectives, although some had served in other verbal learning experiments. The *S*s were assigned to lists in rotation as they appeared at the laboratory.

---

[1] Now at Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

[2] Now at Pomona College, Claremont, California.

The *S*s learned by the anticipation method. All *S*s were given standard paired-associate learning instructions. In addition, *S*s learning the intermittent ORM lists were warned that the response members of the pairs would be omitted a certain proportion of the time and that they should do their best in spite of this handicap.

The lists were presented on a Patterson memory drum at a 2:2-sec. rate. The stimulus term was presented for 2 sec. while a shutter covered the response term. Then the shutter lifted and both the stimulus and the response (or a stimulus and a blank on non-ORM trials) remained in the aperture for an additional 2 sec. The intertrial interval was 4 sec. The criterion for learning and relearning was one perfect recitation of the list. All relearning was with the 100% ORM list. The *S*s reported to the laboratory at the same time on two successive days. Acquisition took place on the first day and relearning on the second day. Hence the average interval between the end of acquisition and the beginning of relearning was approximately 23.5 hr.

A total of 24 *S*s had to be eliminated from the experiment for various reasons. Elimination rate was not differential with respect to conditions, except in the case of failure to understand the instructions for learning. The seven *S*s dropped for this reason were all in either the 20%, 40%, or 60% ORM groups. Therefore, if ability to understand instructions is highly correlated with ability to learn paired adjectives, then the sample of *S*s in these groups may be slightly superior to the *S*s in the 80% and 100% ORM groups.

Statistical analysis.—Analysis of variance was used when the scores under consideration met the requisite assumptions. The *F*s have 4 and 245 *df* in each comparison, with an *F* of 2.37 required for significance at the 5% level. If either heterogeneity of variances or indications of nonnormality were present, then the median test (Dixon & Massey, 1957, p. 296) was used. The χ^2 in this test has 4 *df* in each comparison, and a χ^2 of 9.49 is required for significance at the 5% level.

RESULTS

TABLE 1

COMPARISON OF GOSS, MORGAN, AND GOLIN RESULTS WITH SCHULZ AND RUNQUIST RESULTS

Study	Percentage ORM	Means	
		Trials	Correct Responses
Goss, Morgan, & Golin ($N = 20$)	100	24.8	59.4
	75	34.3	70.2
	50	40.6	94.0
	25	40.5	102.0
Schulz & Runquist ($N = 50$)	100	17.5*	93.2
	80	18.7	102.5
	60	16.8	91.0
	40	21.8	104.5
	20	22.7	113.7

* σ_m based on within groups variance estimate was, for trials =1.22, and for correct responses =5.62. Corresponding *F*s were 4.07 ($P < .01$) and 2.68 ($P < .05$).

Acquisition.—The present results, for measures identical to those reported by Goss et al. (1959), are shown in Table 1.[3] Their results for 8-item syllable lists learned to a $\frac{7}{8}$ criterion at a 2:2-sec. rate of presentation are also presented in Table 1 to facilitate the comparison of the two studies. For the present study, it may be seen from Table 1 that the mean number of trials and correct responses in learning to criterion generally decreased as percentage ORM increased. The "deviant" performance of the 60% ORM group is the only exception to an otherwise monotonic relationship for both measures. The differences among these means are statistically reliable (for trials $F = 4.07$, $df = 4/245$, $P < .01$; for correct responses $F = 2.68$, $P < .05$). It is also apparent from Table 1 that the present results agree well with those obtained by Goss et al., though the differences between the means in the present study are generally of smaller magnitude.

---

[3] A third measure, mean reinforced correct responses, was reported graphically by Goss et al., (1959). Since these means are readily estimated from the data in Table 1 by multiplying the mean number of correct responses for a given group by the proportion of ORM it received, they are not reported here. In accord with Goss' finding, the present mean number of reinforced correct responses increased as percentage ORM increased.

It was our impression, while running *S*s, that *S*s in the 20% ORM group seemed less likely than 100% ORM *S*s to fail to anticipate an item after once having anticipated it correctly. A score designed to show whether our impression was, in fact, a valid one was computed for all *S*s. This score, hereafter referred to as the Correct/Opportunity ratio after first correct, was computed as follows. The *S*'s data sheet was inspected to determine when the response member of a given pair of items was first anticipated correctly. Then the number of subsequent opportunities *S* had to respond to the stimulus member of that pair were noted as well as the number of times he responded correctly. This was done for all 10 pairs on the list. Next, the total number correct was divided by the total number of opportunities. The Correct/Opportunity ratio after first correct is the result of this division. Thus, the less frequently *S* misses items after anticipating them correctly once, the higher his score. The results of this analysis are shown in Fig. 1. Inspection of Fig. 1 reveals that performance in terms of mean Correct/Opportunity ratio after first correct was a monotonic inverse function of percentage ORM. With 20% ORM there were, on the average, 6% more correct responses per opportunity than with 100% ORM. Statistically, the differences among the means of Fig. 1 may be considered reliable ($\chi^2 = 13.8$; $P < .01$). The standard error of these means ranged from .010 to .014.

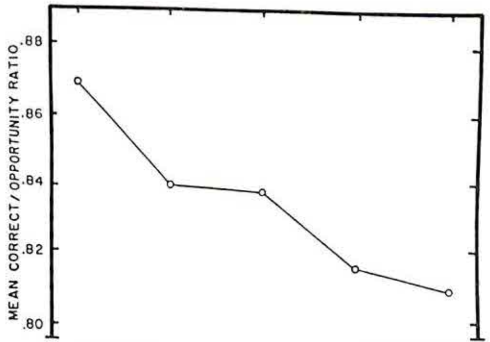


FIG 1. Mean Correct/Opportunity ratios after first correct as a function of percentage of occurrence of response members.

The mean number of overt errors per trial during acquisition was directly related to percentage ORM. The respective means for 100%, 80%, 60%, 40%, and 20% ORM were 1.09, 1.03, .87, .69, and .58. The differences among these means meet the requirements for statistical significance ($F = 6.22$; $P < .01$). The standard error of these means is .085.

Retention.—The mean number of correct responses on the recall trial did not vary as a function of percentage ORM. The means were 6.38, 6.84, 6.58, 6.50, and 6.62 for 100%, 80%, 60%, 40%, and 20% ORM, respectively. The σ_m ranged from .24 to .26. Clearly, the null-hypothesis could not be held suspect ($\chi^2 = 1.92$; $P > .80$). Adjustment for possible differences in response strength at the end of acquisition were made via probability analysis (Underwood, 1954), but the conclusion that percentage ORM did not affect retention remained unaltered. The relearning data mirrored the recall data and therefore need not be presented.

DISCUSSION

Acquisition.—The fact that the present differences in performance as a function of percentage ORM were of much smaller magnitude than those obtained by Goss et al. (1959) can most probably be attributed to the difference in the level of meaningfulness of the learning material employed in the two studies. This seems especially likely because meaningfulness is known to be a potent source of variance in verbal learning (Underwood & Schulz, 1960). Therefore, it appears reasonable to infer from these studies that the effectiveness of variation in percentage ORM may be inversely related to the meaningfulness of the material to be learned.

The general failure to find more marked decrements in performance with reduced ORM surprised us, as it did Goss et al. The *S*s in the present 20% ORM group saw the response members of the list an average of 4.5 times in learning to criterion as contrasted with 17.5 exposures of each response member with 100% ORM. Goss has pointed out some factors which may have counteracted the effect of reduction in ORM. The present results seem pertinent to two of these factors; namely, amount of opportunity for rehearsal and amount of response competition per trial.

Consideration of the inverse relationship between percentage ORM and Correct/Opportunity ratio after first correct (cf. Fig. 1) in conjunction with the data on trials to learn to criterion suggests that the detrimental effects of reduced ORM reflected by the latter measure must have been confined to the relatively earlier stages of learning which primarily involved the accomplishment of the first correct anticipation of a given response member, since maintenance of the response thereafter proceeds better with intermittent ORM. The differential opportunity for rehearsal is also greatest during the early stages of learning and decreases steadily as *S* attempts to anticipate more and more response members in the low ORM condition. Therefore, rehearsal is undoubtedly an important factor in the attenuation of the effectiveness of reduced ORM.

Finally, if the occurrence of overt errors can be taken as an index of response competition, then the direct relationship between frequency of overt errors per trial and percentage ORM found in the present study suggests that response competition does in fact increase as percentage ORM increases. Hence, the second important source of compensation for the reduction of ORM must be associated with reduced response competition throughout learning under conditions of intermittent ORM. Unfortunately, the present data fail to provide any clues as to a possible mechanism by means of which intermittent ORM could reduce competition.

Retention.—The retention results unequivocally failed to support the hypothesis that intermittent ORM increases resistance to forgetting under the conditions of the present experiment. Perhaps the present experiment was simply a poor one for a test of the hypothesis. A better test might be provided by an experiment in which the traditional A-B, A-C retroaction paradigm was utilized to study amount of retroactive inhibition as a function of percentage ORM in original learning.

SUMMARY

The effect of variation in percentage of occurrence (20%, 40%, 60%, 80%, 100%) of response members (ORM) on the learning and retention of paired-adjective lists was studied. A total of 250 *S*s, 50 per list, learned the same list under the five ORM conditions to a criterion of one perfect recitation. Retention was measured 23.5 hr. after acquisition. The following results were obtained:

1. The mean number of trials and correct responses required to reach the criterion decreased as percentage ORM increased.
2. Performance subsequent to the first correct anticipation of a given pair was an inverse function of percentage ORM. That is the ability to maintain a correct association, once made, was greater with intermittent than with continuous ORM.
3. The mean number of overt errors per trial increased as percentage ORM increased.
4. Mean number of correct responses on the recall trial was unaffected by variation in percentage ORM during acquisition.

REFERENCES

DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis.* (2nd ed.) New York: McGraw-Hill, 1957.

GOSS, A. E., MORGAN, C. H., & GOLIN, S. J. Paired-associates learning as a function of percentage occurrence of response members (reinforcement). *J. exp. Psychol.*, 1959, **57,** 96–104.

MELTON, A. W., & IRWIN, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53,** 173–203.

MELTON, A. W., & VON LACKUM, W. J. Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54,** 157–173.

UNDERWOOD, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38,** 29–38.

UNDERWOOD, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42,** 153–161.

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51,** 276–282.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received July 1, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

A TEST FOR INTERACTION OF DELAY OF KNOWLEDGE OF RESULTS AND TWO TYPES OF INTERPOLATED ACTIVITY[1]

EDWARD A. BILODEAU AND FRANCIS J. RYAN

Tulane University

Bilodeau and Bilodeau (1958) recently published results from five studies on delay of knowledge of results (KR) in motor learning tasks. In none of these five studies was delay a significant variable. Delays by seconds, hour, day, and week produced only null results in lever-pulling and knob-turning tasks. Null results were not surprising, for the responses learned were relatively simple and discrete, no special interfering activities were interpolated, and potentially confounding temporal variables were carefully controlled, i.e., intertrial interval and post-KR interval.

Moreover, it was shown that these results were well supported by the literature. There was, however, one outstanding exception that proved difficult to explain without introducing petty criticisms. This was the study by Greenspoon and Foreman (1956) that not only produced significant results but also yielded a handsome set of delay curves with 0-, 10-, 20-, and 30-sec. parameters. Greenspoon and Foreman's conclusion—"The results of this experiment demonstrate the efficacy of immediate knowledge of results in the learning of a motor task,"—was quite at variance with that of Bilodeau and Bilodeau. It was judged that by far the best way to deal with the conclusion was to hypothesize that the findings could not be replicated.

The essence of Greenspoon and Foreman's method was as follows. The *S*s were taught to draw 3-in. straight lines while blindfolded. KR was "long," "short," or "right." There were eight *S*s in a group; each *S* drew 50 lines at 30-sec. intervals. To Greenspoon and Foreman, the critical part of the procedure was that after drawing a line *S* retained his grip of the pencil, kept hand and arm free of the table, and allowed *E* to return the hand to the starting point at the end of the 30-sec. intervals. The foregoing, called "hand-maintaining activity," required *S* to keep his arm off the table for the whole 25-min. testing session and must have been quite tiring.

In an earlier study by Saltzman, Kanfer, and Greenspoon (1955) in which delay of KR had no effect on line drawing, the procedure was different in two respects. First, *S* rested his elbow on the table and briefly raised his pencil when the line was finished. (Again, *E* returned *S*'s hand to the starting point.) Second, the point at which the response originated was only approximately the same from trial to trial. Greenspoon and Foreman allow the possibility that either procedural difference may account for the discrepancy of outcome. As they put it, (*a*) delayed information may differentially reinforce hand-maintaining activity so that longer delays of information are less effective than shorter delays, and (*b*) Saltzman's et al. slightly variable starting point made enough of a stimulus change from trial to trial to reduce the effects of delay.

If we were convinced of the reliability of the above differences in outcome, we might offer arguments about the interpretation, but our basic disagreement is over fact, not interpretation. It is difficult to persuade us, with a single experiment, that such procedural differences as described could make delay of KR effective. The major objective of the

---

[1] This work was supported by a grant from the National Science Foundation.

present study, thus, was to replicate two of their groups. Also tested were two groups of *S*s who kept the hand at rest between responses, with the starting point controlled.

METHOD

Subjects.—The *S*s were 48 undergraduate men and women of Tulane University. Men and women were assigned in equal number to each of four groups.

Apparatus.—The apparatus was patterned after that of Greenspoon and Foreman: a continuous strip of paper rolled over a piece of plate glass fastened to a laboratory table. The paper slid beneath a carpenter's square. Thus, the glass formed a smooth writing surface beneath the paper and the square served as a straight edge for *S*'s pen. The inner corner of the square was the starting point for the pen and, of course, the rulings on the square permitted *E* to measure the length of line as it was drawn. The apparatus was covered with a black cloth except when *S* was blindfolded.

A Lindy pen was used instead of a mechanical pencil. This substitution improved the riding qualities of the instrument along the square. Furthermore, the properties of the writing point were less variable and the point could not be broken. Adhesive tape was wrapped around the pen from 2 to 2½ in. from the point. The tape made a nonslip grip during the hand-maintaining procedures. The tape also served the useful purpose of standardizing the grip between *S*s and for trials within *S*s.

Procedure.—A 2 × 2 factorial combination of delay of KR and maintenance was used. KR was administered almost immediately after a response (I) or was delayed by 20 sec. (D). The hand and arm were maintained (Hold) or they were not (Rest).

The *S*s, blindfolded, had 50 trials. When the first 35 trials were under Treatment I, the last 15 were D trials. When the first 35 trials were under Treatment D, the last 15 were I trials.

The basic sequence of events for all groups was a ready signal, a "draw" signal 5 sec. later, and then KR followed in turn by another ready signal. The draw signals occurred every 30 sec. All four groups were kept to a 30-sec. intertrial interval; the post-KR periods were approximately 30 and 10 sec. long for the I and D treatments, respectively.

KR was "long," "right," and "short." The *E* said "right" when the line was within .25 in. of the 3.00-in. goal. However, *E* recorded *S*'s response error (line length less 3.00 in.) irrespective of KR.

The following practices define hand maintaining and the Rest and Hold treatments.

Treatment rest.—Upon completing his line, *S* returned his hand and arm to his lap. The *S* held the pen, but was otherwise relatively relaxed. "Hand" spoken by *E* was the signal to present hand and pen to *E* who then guided them to the starting point. "Hand" preceded "Draw" by 5 sec.

Treatment hold.—Upon completing his line, *S* initiated no further responses. He maintained his grip of the pen, kept the pen point where he had stopped it, and kept his arm off the table (forearm parallel to the table top). "Hand" spoken by *E* was the signal that *E* was just about to grasp and to guide *S*'s hand to the starting point. Again, the "Hand" signal preceded the "Draw" signal by 5 sec. Treatment Hold was unpleasant, for *S* had to hold his arm out for about half an hour. Nevertheless, all *S*s followed directions.

Design compared with Greenspoon and Foreman.—In all essential respects, Groups I-Hold and D-Hold are replicates of the groups with 0- and 20-sec. delays in the study by Greenspoon and Foreman. One immaterial difference is that our last 15 trials are transfer shift trials. This modification was possible because the differences between their groups were produced during the early trials and not during the late trials.

Had it been possible or desirable, we should, of course, have repeated the condition with the longest delay, i.e., 30 sec. It would seem nearly impossible to have a 30-sec. interresponse interval and delay KR by 30 sec. If the administration were possible, it would be generally undesirable for *S* would have 0 sec. to respond to his feedback.

Greenspoon and Foreman reported the results in terms of right responses. Right responses and the more analytic measure, magnitude of error, will be reported here.

RESULTS

The data were first examined for the mean number of right responses in successive blocks of five trials, an analysis identical with that of Greenspoon and Foreman.[2] There were no consistent differences between groups, there being several changes in rank over blocks of trials. An important question is whether the present results are of the same order of magnitude as those of Greenspoon and Foreman. They obtained mean *rights* of 27.38 and 15.00 for Groups I and D, respectively. Projecting our 35-trial results to a base of 50 trials, we obtain 27.74 and 26.67 for Groups I and D, respectively. There is, thus, remarkable agreement for I and marked disagreement for D. Obviously, our D *S*s outperformed theirs. Our D *S*s were about twice as variable, but our I *S*s were comparable. Moreover, within their own experiment, D *S*s showed the lesser variability. It is our guess, then, that the fundamental contradiction between studies is attributable to their D group which is low in mean and variability.

---

[2] Tables of data have been deposited with the American Documentation Institute. Order Document No. 6301, remitting $1.25

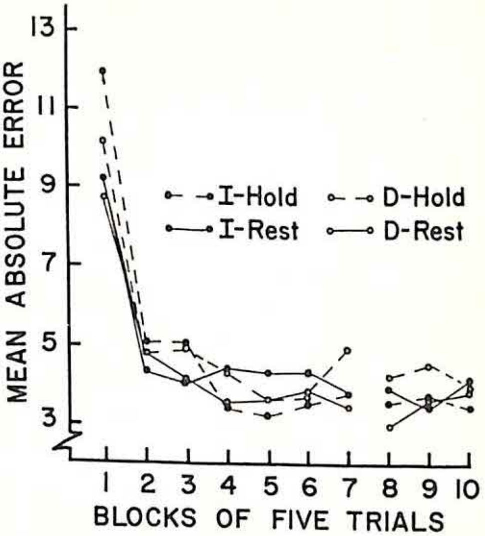


FIG. 1. The effect of delay of knowledge of results and of maintenance. For Blocks 8, 9, and 10, the I groups were switched to Delay and the D groups were switched to Immediate KR.

A more analytic measure of the line-drawing response is the mean absolute error. Means based upon successive blocks of five trials are plotted in Fig. 1. The left side of the figure represents the trials before the shift in delay, the right side the trials after the shift. The only discernable trend is that of learning; no differences appear ascribable to delay of KR or to maintenance.

The means and variances of the totals for *right* and *errors* per *S* over the first 35 trials are given in Table 1. The column means, based upon 24 *S*s, show that the greater number

TABLE 1

MEANS AND VARIANCES FOR THE FIRST 35 TRIALS
(Right represents the number correct; error (1/16 in.) represents the absolute error per *S*)

Maintenance	Delay				Mean	
	I		D			
	Right	Error	Right	Error	Right	Error
Hold	19.42 (12.63)	185.17 (2088.15)	18.67 (18.79)	184.33 (2596.42)	19.04	184.75
Rest	18.42 (13.72)	174.83 (648.70)	19.50 (11.54)	162.75 (1053.11)	18.96	168.79
Mean	18.92	180.00	19.08	173.54	19.00	176.77

---

for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies.

TABLE 2

MEANS AND VARIANCES FOR THE LAST 15 TRIALS
(Right represents the number correct; error (1/16 in.) represents the absolute error per *S*. Group I has been switched to Delay and Group D has been switched to Immediate KR)

Maintenance	Former Delay				Mean	
	I		D			
	Right	Error	Right	Error	Right	Error
Hold	10.25 (1.66)	54.75 (148.93)	8.58 (9.36)	64.50 (546.82)	9.42	59.62
Rest	9.17 (6.52)	58.25 (202.20)	9.75 (5.30)	52.92 (267.17)	9.46	55.58
Mean	9.71	56.50	9.16	58.71	9.44	57.60

of rights and the lesser error were obtained by Treatment D. Neither the Delay nor Maintenance variable, however, exerts much influence for all of the entries are exceedingly similar. This was amply confirmed by analysis of variance of the two factorials in Table 1. Most of the *F*s were less than 1.00 and none approached the .05 level.

The right responses and the errors for Trials 36–50 are shown in Table 2. These are the transfer-shift trials; Treatment I actually received delayed information and Treatment D received immediate information. Once again all of the entries are very similar. In terms of numerical differences, both measures of response favor the treatment with former immediate KR and present delay of KR. The *F* tests, however, failed by a wide margin to reach the .05 level.

There are eight opportunities to compare delay with immediate information in the cells of Tables 1 and 2. In five of the eight, the better performance is associated with the delay treatment. These comparisons are mentioned in order to make a minor point—the difficulty in maintaining that the better treatment is I. The major point is that there is no source of significant variance in Tables 1 and 2.

DISCUSSION

There was no evidence of special efficacy of immediate KR in the present study. Greenspoon and Foreman give evidence of such efficacy when the hand was maintained. They suggested that the positive result might be obtained when the hand was maintained, but not when the hand was not maintained. We have found no evidence that maintenance, nonmaintenance interacts with delay of KR, and no evidence that delay of KR impairs the learning process. Further, the maintenance variable itself did not produce significant differences.

The effects of delay of KR upon the learning process have been recently discussed by Bilodeau and Bilodeau (1958). Maintenance is another matter. It may or may not interact with delay of KR, depending upon what we mean by the word.

Meanings of maintenance.—If *S* suspended activities after drawing a line, the traces of the response would be undisturbed. A complete suspension has never been achieved. It is typical procedure for *S* to initiate and execute a number of responses other than line drawing. Some of these involve the hand and arm and their movement to the table top, to the starting point, and so on. These extra responses are included within nonmaintenance procedures. As used here, maintenance also refers to certain observable responses of the hand and arm, e.g., the pencil- and arm-holding actions. Maintenance, as Greenspoon and Foreman recognized, was not the opposite of nonmaintenance. Neither the Rest nor Hold techniques used here produced or even approximated a state of suspended activity.

Since the Thorndikian line-drawing task is not especially suited to achieving anything but a moderate degree of control over *S*'s extra responses, we prefer the Rest treatment as standard, neutral procedure. At least at present, *S* must do something with his arm and hand before and after KR, and the hand must be positioned for the beginning of the drawing response. Bilodeau and Bilodeau have suggested that dropping the arm to one's side after responding and then engaging in normal waking activities offer just about as standardizing (and neutral) a situation as we are likely to find for some time. The value of normal activity as a standard procedure is underscored when pre- and post-KR delays are of the order of hours, days, and even years. Maintenance at any level (limb or trace) would seem out of the question with such intervals.

We allowed that both Rest and Hold activities represent potential interfering tasks, though not differential ones. Our null results do not permit concluding that level of interference varies from Rest to Hold, from I to D, and in the interaction of these two variables. But the extent to which the outcome of the present study is restricted by our range of values is most pertinent. As far as delay of KR is concerned, the range of variation seems adequate, Bilodeau and Bilodeau (1958) having shown that delays of KR up to week-long intervals had no adverse effects upon learning. As to the separation between Rest and Hold we are not quite as certain. First, by *S*s introspections, Hold is a lot more tiring than Rest (see the section below). Second, our critique of the procedures Rest and Hold shows that neither is a condition of suspended activity. Third, the Rest-Hold differences in the results were null. Thus, there are two good arguments that Rest is not especially different from Hold.

Someone will no doubt suggest that the second argument above shows the need for a control group which is completely free of interpolated activity. This type of control is not yet technically feasible as is made clear in the debates over the techniques of freezing, hypnosis, and drugs in testing retroactive inhibition theory.

Effects of Rest and Hold treatments.—There was no over-all difference between the two conditions of maintenance. This is a puzzle if nonmaintenance (Rest) implies a condition of relative rest between line-drawing responses, and maintenance (Hold) implies a condition of work. Having equated maintenance with a species of massed practice, and nonmaintenance with spaced practice, we might expect performance to favor nonmaintenance. It might be that a condition of massed activity plus less interference has been placed in balance against spaced activity plus greater interference. Though this is a possibility, our preferred interpretation of the null result is that (*a*) the interferences are essentially equal, and (*b*) the differential fatigue exists but is irrelevant *to the accuracy* of line drawing. By analogy, the overtired lathe worker can produce sufficiently accurate pieces of work. Such pieces can be produced despite protestations of fatigue. In the laboratory, the ancient ergographologists often succeeded in producing a good deal of fatigue though little finger-decrement. Similarly, while Hold *S*s may be more fatigued than the Rest *S*s, this does not imply that they cannot draw lines as well. It does imply that there is *some* activity which cannot be performed as well. This "some" activity need not necessarily be the same activity which produced the intervening state. Under the circumstances where the state does depress the activity that produced the state, the decrement may be in one measure of response and not in another.

SUMMARY

This experiment replicates two groups of a line-drawing study by Greenspoon and Foreman, i.e., 0- and 20-sec. delay of knowledge of results. Their data gave evidence of the efficacy of immediate knowledge of results. The present experiment gave null results, the outcome being entirely consistent with a number of earlier studies.

The *S*s of the two groups mentioned above followed a "hand maintaining" procedure in that the drawing hand and arm were kept off the table between responses, and in that *E* positioned *S*'s hand for the start of each response. Two other groups, also 0- and 20-sec. delay, did not maintain the responding limb, but returned it to the lap and rested between responses. Neither procedure led to differences between immediate and delayed information.

The maintenance procedure used here was very tiring. Nevertheless, there was no difference between any maintenance and nonmaintenance groups. This was interpreted to mean that fatigue is selective and need not impair the accuracy of all types of responses even over an extended number of trials.

Possible maintenance procedures and how they might influence the trace of the line-drawing response are discussed.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD., Variation of temporal intervals among critical events in five studies ot knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

GREENSPOON, J., & FOREMAN, S. Effect of delay of knowledge of results on learning a motor task. *J. exp. Psychol.*, 1956, **51**, 226–228.

SALTZMAN, I. J., KANFER, F. H., & GREENSPOON, J. Delay of reward and human motor learning. *Psychol. Rep.*, 1955, **5**, 139–142.

(Received July 1, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

EFFECTS OF SITUATIONAL ANXIETY AT DIFFERENT STAGES OF PRACTICE[1]

JOAN L. BARDACH[2]

New York University

The hypothesis investigated in the present experiment is that situational anxiety induced early in practice will have a more unfavorable effect upon performance than anxiety induced late in practice. Insofar as anxiety is viewed as a drive that energizes all activities that may be in progress, the stage in practice at which it is introduced should affect performance. Early in practice, the strength of the correct responses relative to that of the incorrect responses is less than it is later in practice. Indeed, early in practice, especially with difficult tasks, the incorrect responses may be the stronger, and introduction of anxiety therefore may be detrimental to learning and performance. Later in practice, however, the strength of the correct responses may exceed that of the incorrect ones, so that the introduction of anxiety should facilitate performance, or at least be less detrimental than earlier introduction would be.

If we consider anxiety as a class of responses, we can make the same prediction. The responses associated with anxiety possess a certain degree of strength. If these responses are of such a nature that they interfere with the correct responses, the observed effects of this interference will depend upon the relative strength of these two kinds of reactions. In an early stage of practice, the strength of the correct responses is less than it is in a later stage. Hence the amount of impairment of performance produced by interfering responses associated with anxiety should be less in later stages of practice. In making this prediction, it is assumed that the strength of the anxiety reactions does not increase markedly during learning.

METHOD

Subjects.—The *S*s for this experiment were volunteer, undergraduate men from New York University. Only male *S*s were used in order to eliminate sex differences in susceptibility to electric shock.

Learning material.—The learning material was a list of 10 nonsense syllables, selected from Glaze's list (1928) and of 26.67% association value. Nonsense syllables of low association value were used because previous experimental results indicated that such syllables present a learning task sufficiently difficult for learning to be impaired either by high anxiety, as measured psychometrically (Taylor, 1953), or by situational anxiety induced by unavoidable shock.

Apparatus.—Each nonsense syllable was exposed for 2 sec. on a Patterson memory drum, Model 1B. The intertrial interval was 20 sec. Shock was administered by means of a Porter inductorium adjusted for tetanizing shock; i.e., the vibrator make-break mechanism was used. The primary coil and make-break vibrator were in series with a telegraph key and one No. 6 dry cell.

The electrodes were pieces of aluminum

---

[1] Based on a dissertation submitted in partial fulfillment of the requirement for the PhD degree at New York University. The author gratefully thanks Leland Crafts, Research Advisor, whose constructive ideas and supportive attitudes contributed so materially to this research. The author also thanks Ralph Gilbert and Robert Silverman, members of her Thesis Committee, for their helpful comments in which problems in method and theory were anticipated.

[2] Now at the Institute of Physical Medicine and Rehabilitation.

mounted side by side on a wooden block, against which *S*'s fingers were strapped. Leads to the electrodes were taken from the secondary terminals of the induction coil.

Procedure.—The *S* was asked to anticipate each syllable by spelling it. In accordance with the procedure of Silverman and Blitz (1956), all *S*s were given a series of preliminary training trials in which they were required to learn a list of six simple three-letter words that were presented in the same manner as the experimental task. Following the preliminary training, each *S* learned the experimental list of nonsense syllables to the criterion of two consecutive errorless trials.

Four groups of *S*s, 17 per group, were used. At the prescribed stage of practice, *S* in an experimental group was suddenly told: "From now on you will receive electric shocks at any time." While these instructions were being given, electrodes were attached to *S*'s fingers. The total time required for the verbal instructions and for attaching the electrodes was 30 sec. At the conclusion of the trial following this procedure, one shock actually was given. However, no shock was given at any later time, although presumably *S* continued to expect one.

Shock was administered in the experimental groups either after *S* had learned 30% of the list, i.e., had anticipated three syllables correctly (Group E3) or after *S* had learned 70% of the list, i.e., had anticipated seven syllables correctly (Group E7). Group C3, the control for Group E3, was equated with the latter in terms of the number of trials to reach the criterion of 30% correct anticipations on the experimental list. Group C7, the control for Group E7, was equated with the latter in terms of the number of trials to reach criterion of 70% correct anticipations on the experimental list. Since the two experimental groups had an interval of 30 sec. following the learning of 30% or 70% of the list, respectively, their control groups had an interval of the same length following the achievement of the same criteria. In order to prevent rehearsal of the rote-memory task by the control *S*s, they were asked simple questions, such as age, college major, etc., during this 30-sec. interval.

After completion of the learning task, each experimental *S* filled out a questionnaire designed to reveal his reactions to shock-warning and to shock-administration. At the conclusion of the experiment, all *S*s were given an explanation of the experiment and attempts were made to allay whatever anxiety might have been aroused.

RESULTS

Results of the analysis of variance, using mean total trials to learn to criterion, are shown in Table 1. These data indicate no statistically reliable differences in performance among the groups as a whole, nor between the groups that were given shock and those that were not, nor between those groups that were interrupted early in learning and those that were interrupted late. Further study of the interaction between time of interruption and shock—no shock again revealed no differences. Comparisons of trials or errors to criterion for the various groups by t tests (matched groups formula used where appropriate; two-tailed test of significance) revealed only one significant ($P < .05$) difference between the means. Between Groups E7 and C7, the t was 2.68 ($P = .02$, $df = 16$). This finding indicates that it is the late introduction of anxiety that results in impaired performance, rather than its early introduction as was the theoretical expectation. This result does not conform to predictions based on a view of anxiety either as a drive or as a class of interfering responses.

However, qualitative data noted during the running of the experiment, as well as the results of the question-

TABLE 1

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION

Source	*df*	*MS*	*F*
Between pairs	33		
B (7 vs. 3)	1	187.78	1.59
Error (b)	32	117.77	
Within pairs	34		
A (E vs. C)	1	101.31	.90
AB	1	106.25	.94
Error (w)	32	112.44	
Total	67	115.85	

TABLE 2

NUMBER OF TRIALS TO LEARN TO CRITERION

Groups Compared	N	Mean	SD	t
E3 Compat.	6	24	10	2.61*
C3 Compat.	6	39	13	
E3 Incompat.	11	34	9	2.32*
C3 Incompat.	11	28	13	
E7 Compat.	6	29	12	.76
C7 Compat.	6	25	10	
E7 Incompat.	11	31	9	2.71*
C7 Incompat.	11	23	9	
E3 Compat.	6	24	10	2.51*
E3 Incompat.	11	34	9	
E7 Compat.	6	29	12	1.38
E7 Incompat.	11	31	9	
E3 Compat.	6	24	10	.74
E7 Compat.	6	29	12	
E3 Incompat.	11	34	9	1.04
E7 Incompat.	11	31	9	

* $P = .05$.

naire and a brief terminal interview given the experimental *S*s, suggested that the attitudes induced by the introduction of shock were not the same for every *S*. Failure to take this fact into account in the statistical treatment of the data may explain the negative results obtained.

Therefore, each experimental group was divided into two subgroups, based upon revealed differences in the attitudes induced by shock introduction. One subgroup was composed of *S*s who claimed they were not disturbed by the shock or who believed, despite the instructions, that they would be shocked for *errors* and therefore may have been spurred to greater effort to learn, or of those who felt that by diligent effort they could complete the experiment quickly and thus avoid further shock, or who did not continue to expect shock throughout most of the learning. This group is called the *Compatible* group because the attitudes induced by the experimental conditions seemed to be compatible with the requirements of the learning task. The designation E3 Compat. and E7 Compat. refer to the experimental *S*s who were given shock early or late in the learning, respectively, and in whom compatible attitudes were induced. Groups C3 Compat. and C7 Compat. are comprised of those control *S*s who matched the experimental *S*s in performance before the introduction of the experimental conditions.

The second subgroup into which the experimental groups were divided consisted of those *S*s who were disturbed by the shock, or who felt that shock would be administered randomly, or who continued to expect shock throughout most of the learning. This group is called *Incompatible* because the attitudes induced by the experimental conditions seemed to be incompatible with the requirements of the learning task. The designation E3 Incompat. refers to the experimental *S*s who were given shock early in the learning and in whom incompatible attitudes were induced; C3 Incompat. refers to their control group. E7 Incompat. and C7 Incompat. are the corresponding designations for the experimental group that was given shock late in the learning and its control.

Table 2 shows comparisons of performance when the compatibility or incompatibility of responses associated with the secondary anxiety drive are taken into account. As the table shows, *S*s in the compatible group who received shock early in practice, learned significantly faster than their controls, but *S*s in the compatible group who received shock late in practice did not.

Performance data for the incompatible groups are not simply the reverse of those obtained for the compatible groups. Table 2 shows that the experimental *S*s in whom incompatible responses were induced learned significantly slower than their controls, regardless of whether shock was introduced early or late in practice. The evidence is that Group E3 Incompat. did not take significantly more trials to learn to criterion than did Group E7 Incompat. However, because of the small *N* of 11 in each group, further research is necessary to establish this point definitely.

In order to analyze the course of the effect, or of the immediate and the subsequent effects, of shock introduction, apart from its total over-all effect upon *N* of trials (and of errors) to complete the learning, various learning curves[3] were drawn. Inspection of the curves showing the mean number of correct responses for each trial for the subgroups who received shock *early* in the learning discloses that, after the introduction of the experimental conditions, the performance of the experimental groups in relation to their controls is reversed. After shock, Group E3 Incompat. acquires additional syllables *more slowly* than its control, at least for the subsequent 12 trials, but, after shock, Group E3 Compat. acquires additional syllables *faster* than its control.

Curves were also drawn to show the mean number of trials necessary to learn 1, 2, 3 . . . 10 syllables, and to criterion. Inspection of these curves for the subgroups who received shock *early* in the learning discloses that, after introduction of the experimental conditions, the performance of the experimental groups in relation to their controls is again reversed. Group E3 Compat., throughout the remainder of the learning, acquired each syllable subsequent to shock in *fewer* trials than its control, but Group E3 Incompat., throughout the remainder of the learning, acquired each syllable subsequent to shock in *more* trials than its control.

The comparable curves for the subgroups who received shock *late* in the learning show that the experimental conditions had little effect on performance of those individuals in whom compatible attitudes were induced. It may be that there was insufficient time for these attitudes to take effect. However, these data show that when incompatible attitudes are induced even late in the learning, the acquisition of subsequent syllables is impaired.

DISCUSSION

The data for the groups as a whole before their division into subgroups in accordance with the compatibility or incompatibility of the induced responses indicate that only late introduction of anxiety impaired performance. If these findings had been the only ones available, the results of the experiment would have been negative, and moreover, would have been difficult to explain in terms of any current theoretical position.

However, when the experimental groups are classified according to the type of induced response associated with the introduction of shock, differences in performance appear that seem more understandable. Now, if the responses to anxiety are compatible with the learning task, early introduction of anxiety is facilitating, although late introduction remains ineffective. If the responses are incompatible, both early and late introduction of anxiety is impairing, although no difference in the degree of impairment is demonstrated. These results support the rather obvious notion that if anxiety is associated with incompatible (interfering) responses, it will impair learning and performance, but they also indicate that if anxiety is associated with compatible responses, it may significantly facilitate learning and performance, even when task difficulty is held constant.

---

[3] These curves can be found in the original dissertation.

However, one problem in the interpretation of the results remains, namely, that when anxiety is introduced late rather than early in practice, even those *S*s in whom compatible responses were evoked were inferior to their controls. A possible explanation of this finding is as follows. In a situation in which compatible responses are induced, a time-lag probably occurs because time is required for the compatible responses associated with the anxiety to produce their potential effects on performance. Table 2 shows that the difference in performance between Groups E7 Compat. and its control is not significant. It may be that the introduction of shock in this experimental group occurred too late in learning for the effects of compatible attitudes to be reflected in performance.

The present results suggest that anxiety can best be conceptualized as having both drive properties and associated responses, a view that is similar to that of Mandler and Sarason (1952). If accurate predictions of performance are to be made, both of these aspects must be taken into account. The data suggest, in addition, that responses associated with the introduction of shock are different for different *S*s, so that accurate predictions require that these individual differences also be considered. However, since the conclusions from the present study are based on *post hoc* considerations, they must be regarded as tentative until cross-validating studies are made.

SUMMARY

An experiment tested the hypothesis that situational anxiety induced early in practice will have a more unfavorable effect upon performance than anxiety induced late in practice. Four groups of *S*s learned a difficult list of nonsense syllables. One group was given unavoidable electric shock after they had learned 30% of the list; another group was given unavoidable shock after they had learned 70% of the list. Each experimental group was matched on preshock performance with a control group.

The data showed that it was late rather than early introduction of situational anxiety that impaired performance, indicating a need for modification of the original hypothesis. When each experimental group was divided into two subgroups in accordance with *S*s' reported attitudes toward shock, it was found that the different responses induced by the shock had significant effects on performance. The results suggest that when shock induces responses that are compatible with the learning task, facilitation of performance may occur, but that when shock induces responses that are incompatible with the learning task, performance is impaired regardless of when in the course of learning shock is introduced.

REFERENCES

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

MANDLER, G., & SARASON, S. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–174.

SILVERMAN, R. E., & BLITZ, B. Learning and two kinds of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 301–303.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received July 2, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

ROLE OF POSTURAL EXPERIENCES IN PROPRIOCEPTIVE PERCEPTION OF VERTICALITY[1]

RICHARD G. PEARSON

Carnegie Institute of Technology

AND GEORGE T. HAUTY[2]

School of Aviation Medicine, USAF

In a previous study (Pearson & Hauty, 1959) concerned with the function of learning in the proprioceptive perception of verticality, it was found that alternation of the direction of tilt on successive trials effected a progressive reduction in the error of perceived verticality. In accounting for this effect, two possibilities were discussed: the learning of additional and unique proprioceptive cues to be gained from alternation and the nullification of effects of sensory adaptation which also could have been provided by this experience.

The present study was designed for the purpose of determining the effects of adaptation imbalance and of increase in the delay at tilt upon the error of perceived verticality. Should these effects attenuate reduction in error, sensory adaptation may be regarded as the predominant factor. If not, then support for proprioceptive learning will have been realized.

METHOD

Apparatus.—The tilting-chair used in this study has been described in detail elsewhere (Pearson, 1957). The seat was adjustable to the sitting height of *S* so that the medial axis at eye level would be approximately at the point of chair rotation. The *S*'s control panel contained adjustable, leather sleeve cuffs to support his forearms at positions of tilt, and a three-position switch providing him with directional control of the chair which, of course, was also under the independent control of *E*. The test room itself was lightproof and airconditioned. Both red and white light were available to *E* at his discretion.

Design.—The experimental variables were of the following nature: (*a*) Alternation, or directional sequence of tilt, within each of the two magnitudes of tilt: A_1—no alternation, in which *all* trials involved the tilting of *S* to his left; A_2—alternation, in which *S* was tilted to the left, either 10° or 30°, on all odd-numbered trials and 10° to the right on *all* even numbered trials. (*b*) Tilt magnitude: T_1—10° left; T_2—30° left. (*c*) Delay at tilt: D_1—no delay; D_2—30 seconds delay; D_3—60 seconds delay.

The *S*s were assigned in equal numbers and randomly to each cell of the $2 \times 2 \times 3$ factorial design. The experimental sample was obtained from a population of volunteer, experimentally naive, basic Air Force trainees. Four *S*s were tested per day until a sample of 96 was attained.

Procedure.—The *S* was secured in the tilting-chair, then was indoctrinated in his task. He was told that the test session would consist of 20 trials. If he had no questions, testing then commenced. The *E* moved *S* to a position of tilt, then, following the delay period (if any), transferred control of the chair to *S* who made his judgment. When satisfied with his position, *S* responded "OK" or "All right." The *E* then recorded the direction and amount of error. Thereafter, *E* proceeded in accordance with the "alternation" condition to which *S* was assigned. Delay at the perceived vertical was 20 sec. in all cases.

---

[1] The contents of this manuscript reflect the personal views of the authors and are not to be construed as a statement of official Air Force policy.

[2] Now at the Civil Aeromedical Research Institute, FAA, Oklahoma City.

TABLE 1

ANALYSIS OF VARIANCE OF ABSOLUTE ERRORS

Source	*df*	*MS*	*F*
Alternation	1	335.48	4.75*
Tilt	1	3613.45	51.16**
Delay	2	20.74	.29
A × T	1	39.81	.56
A × D	2	76.14	1.08
D × T	2	135.37	1.92
A × T × D	2	103.33	1.46
Error (a)	84	70.63	
Measures	9	9.83	2.42*
M × A	9	6.54	1.61
M × T	9	8.66	2.13*
M × D	18	2.62	.64
M × A × T	9	5.17	1.27
M × A × D	18	3.53	.87
M × D × T	18	2.30	.56
M × A × T × D	18	4.87	1.20
Error (b)	756	4.07	
Total	959		

* $P < .05$.
** $P < .01$.

RESULTS

In accordance with the main interests of this paper the statistical analysis was confined to the absolute magnitude of error in perception of the vertical for the odd-numbered trials only. Results of a repeated-measurements analysis of these data are shown in Table 1. Of particular concern to the authors is the finding of significant effects attributable to both the Alternation and Tilt variables. As seen in Fig. 1 the alternation subgroup (10° R), as compared with the "control" subgroup, demonstrates a marked reduction in error beginning at Trial 3, and maintains this superior performance throughout testing. Figure 2 plots the effects of magnitude of tilt upon perception of the vertical, and, as one would expect, the 30°-tilt subgroup has a much larger error of adjustment than the 10°-tilt subgroup. It should be noted at this point that, although both Alternation and Tilt main effects are significant, neither the simple interaction, A × T, nor the second-order interaction, M × A × T, approach significance. The importance of these findings are discussed below.

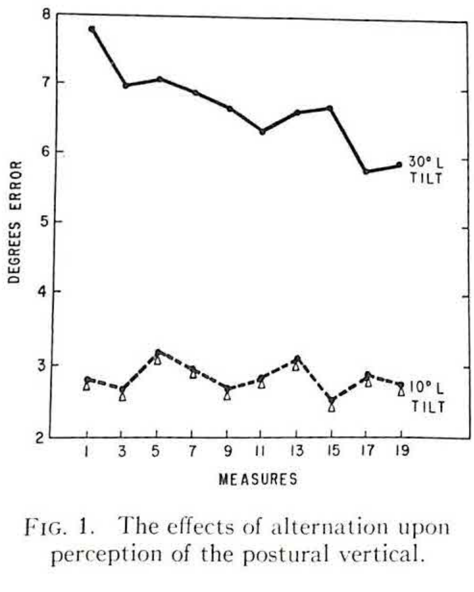


FIG. 1. The effects of alternation upon perception of the postural vertical.

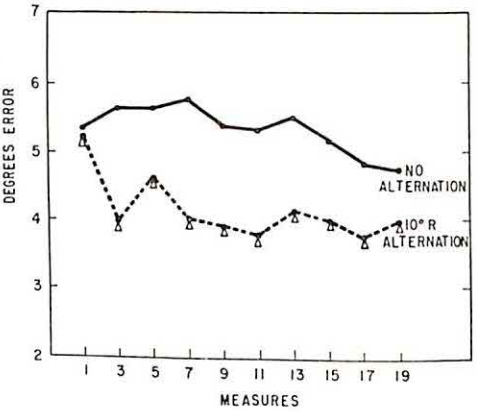


FIG. 2. The effects of magnitude of tilt upon perception of the postural vertical.

The factor of Delay has no significant effect. As shown in Table 1, the Delay *F* ratio is not significant, and, furthermore, none of the interactions involving the Delay variable with the other main variables are significant. The absolute mean errors of adjustment for the 0-, 30-, and 60-sec. delays were 4.64°, 5.05°, and 4.59° respectively.

The two significant *F*s, Measures and M × T, are doubtless attributable to the progressive reduction in error achieved by the 30°-tilt subgroup (see Fig. 2).

DISCUSSION

The primary concern of the present study was with the effect of alternation upon perception of the postural vertical. On odd-numbered trials all *S*s were tilted to the left; on even-numbered trials half of these *S*s were tilted a small amount past the true vertical to the right while the other half were tilted again to the left. It was then demonstrated that perception of the vertical by the alternation subgroup was significantly superior to that of the control subgroup. Further, since the interactions, A × T and M × A × T were found not to be significant, it seems reasonable to assume that adaptation imbalance did not attenuate progressive reduction in error of perceived verticality. Similarly, since neither the F ratios for the Delay variable nor those involving interaction of Delay with the other main variables proved to be significant, one must conclude that increased delay at tilt also did not produce such attenuation. Such findings lead the authors to conclude that proprioceptive learning, rather than sensory adaptation, is the determining factor in progressive reduction in the error of perceived verticality.

The confirmation of our previous finding (Pearson & Hauty, 1959) that "passing through the vertical" is an important variable in perception of postural verticality may have important implications regarding the results of work by Mann, Berthelot-Berry, and Dauterive (1949) and Solley (1956). They employed the practice of moving *S* to the opposite quadrant before moving him to the trial position as a precaution against the possibility that *S* would count himself back to the true vertical. It is now clear that this procedure plays a definite and positive role in reducing errors of perception of verticality. Consequently, it is questionable whether the "true" effects of such variables as delay at tilt and rate of chair movement have been isolated in studies wherein alternation has been employed as a precautionary measure.

In agreement with results reported by Clegg and Dunfield (1954) is the finding that delay did not have a significant effect upon perception of verticality. This confirmation supports a previous conclusion (Pearson & Hauty, 1959); namely, that averaging the results of several consecutive trials does not provide a clear picture of the state of performance in these studies. Performance curves may be expected to reveal that with repeated testing certain variables do not have a significant effect upon performance.

SUMMARY

Ninety-six *S*s were tested on a lateral-tilt chair under darkroom conditions. The task consisted of 20 trials and required *S* to return himself from a position of tilt to the perceived vertical. It was found that neither the procedure of effecting an adaptation imbalance by tilting *S* 30° L, then 10° R, on alternate trials, nor that of increasing the delay at tilt to 30 or 60 sec. attenuated progressive reduction in error of perceived verticality. This would suggest proprioceptive learning, rather than sensory adaptation, to be the factor of principal importance.

REFERENCES

CLEGG, W. C., & DUNFIELD, N. M. Non-visual perception of the postural vertical: II. Lateral plane. *Canad. J. Psychol.*, 1954, **8**, 80–86.

MANN, C. W., BERTHELOT-BERRY, N. H., & DAUTERIVE, H. J., JR. The perception of the vertical: I. Visual and non-labyrinthine cues. *J. exp. Psychol.*, 1949, **39**, 538–547.

Pearson, R. G. The effects of motion-sickness preventives upon orientation in space. USAF Sch. Aviat. Med., November 1957, Rep. No. 58–7.

Pearson, R. G., & Hauty, G. T. Adaptive processes determining proprioceptive perception of verticality. *J. exp. Psychol.*, 1959, **57**, 367–371.

Solley, C. M. Reduction of error with practice in perception of the postural vertical. *J. exp. Psychol.*, 1956, **52**, 329–333.

(Received July 3, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

SUPPLEMENTARY REPORT: THE YERKES-DODSON LAW AND SHIFT IN TASK DIFFICULTY[1]

VICTOR H. DENENBERG AND GEORGE G. KARAS[2]

Purdue University

Broadhurst (1957) has reported a confirmation of the Yerkes-Dodson Law by use of an underwater Y-maze in which motivation was manipulated by the amount of oxygen deprivation prior to the release of *S*, and task difficulty was varied by changing the relative difference in brightness between the two arms of the Y. He also reported that emotional *S*s exhibited a higher drive level than nonemotional *S*s. We used Broadhurst's technique to study the effects of infantile handling since it has been suggested that such handling modifies *S*'s emotionality. We report here a pilot study using a straightway as well as some of the Y-maze data of our control *S*s from the infantile experience study. These groups may be considered to provide a partial replication of Broadhurst's study with different aged *S*s and with somewhat different delay intervals.

Method.—A replica of Broadhurst's maze was used. It was illuminated by two 100-w. bulbs. A black wooden block at the end of the incorrect alley served as the barrier. The correct alley had no barrier. The difference in alley illumination was determined by a Western illumination meter, No. 756, with filters removed. The difference was in the ratio of 1.50.

In a pilot study 5 trials per day were given 21 Harvard Wistar rats (55–61 days old) in a 4-ft. underwater straightway. Time delays of 0, 3, 6, 9, 12, 15, and 18 sec. were used in a 7 × 7 latin square which balanced test days and test sequences. Swimming time was recorded in .01 sec. In the Y-maze study, time intervals of 0, 3, 6, and 12 sec. were used, with 14 *S*s randomly assigned to each condition. Starting at 51 days *S*s were given 5 trials a day for 10 successive days. The time in .01 sec. from the moment that the startbox was opened until *S*'s head broke water was recorded, as was the number of errorless trials per day.

Results.—The pilot study yielded significant differences ($P < .01$) among Days and Delay Conditions in an analysis of variance of swimming times. The plot of time against days is similar in form to Broadhurst's Fig. 1. The relationship between swimming time and delay time is highly irregular, but increased time tends to be associated with longer delay intervals.

The Y-maze *S*s averaged 86.6 errorless trials out of 100. This is similar to Broadhurst's "easy" discrimination task where he found a mean of 86.4 errorless trials out of 100. Analyses of the log time scores and number of errorless trials failed to yield significant differences among the motivational levels. However, a significant Days × Motivation interaction ($P < .01$) was found for the time data. The errorless trial data showed that *S*s had virtually learned the discrimination by the end of Day 5, making 97.3% correct responses during Days 6–10. Thus behavior during Days 1–5 reflects *acquisition* of the response; behavior during Days 6–10 reflects *performance* of the acquired response. The swimming times were separately analysed

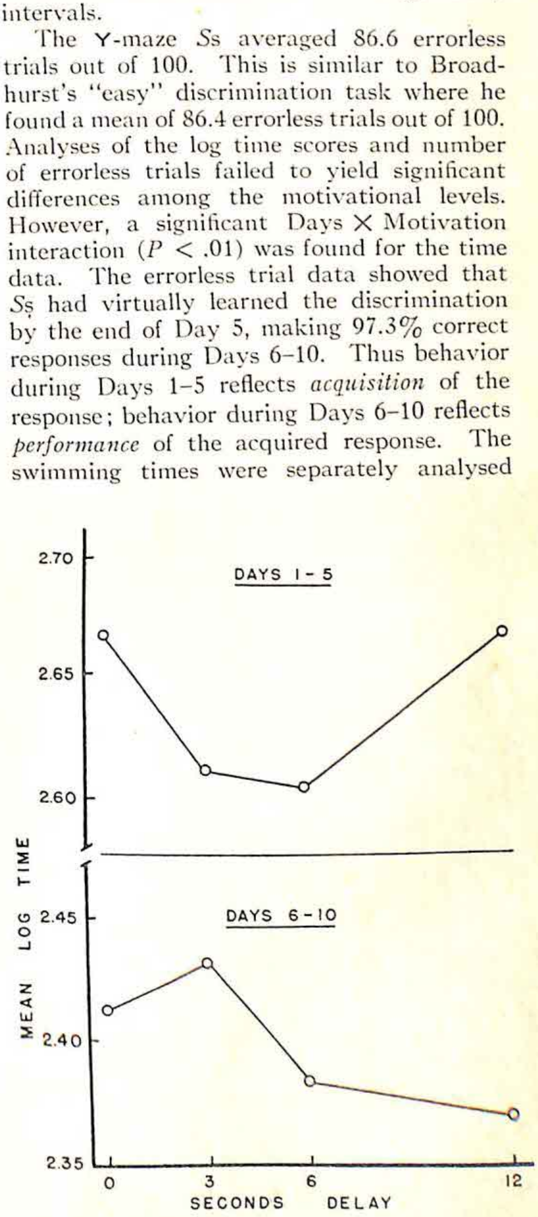


FIG. 1. Swimming time as a function of delay interval for Days 1–5 and 6–10.

---

[1] This study was in part supported by a grant from the Purdue Research Foundation and in part by Research Grant M-1753 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Presently at Department of Psychology, Iowa State University.

for each 5-day period as a function of delay interval. Analyses of these curves (Fig. 1) yielded a significant ($P < .01$) quadratic (curvilinear) effect between swimming time and delay time during Days 1–5 and a significant ($P < .01$) linear function during Days 6–10.

Discussion.—Broadhurst, in his straightway study, tested *S*s first at the 0-sec. delay, then at the 5-sec. delay, etc., thus confounding practice effects and delay intervals.[3] Our data, which separately evaluated these two variables, are not consistent with his, but neither are they consistent with our Y-maze results. It appears that the use of a design in which *S*s are tested under different delay times is not valid because of differential transfer effects.

Our light–dark ratio was approximately 1:50 while Broadhurst's was about 1:300 for his "easy" task, but the two studies report almost the same percentages of correct responses. However, Broadhurst's "moderate" task had a light–dark ratio of 1:60 (83.1% correct responses); and our data should have come closer to approximating the findings of his "moderate" rather than his "easy" task. The failure to do so can probably be accounted for by the ages of *S*s involved (213 or 322 days vs. 51 days) and strain differences.

Even though the main effect of Motivation was not significant, our Y-maze results, like Broadhurst's, confirm the Yerkes–Dodson Law when task difficulty is changed by repeated experience. The task changed markedly in difficulty between Trials 1–5 and 6–10. According to the Yerkes-Dodson Law, optimal motivation should be less during the more difficult learning trials (Trials 1–5) than during the latter trials. This effect was obtained, and suggests that *change in task difficulty* as a function of repeated practice must be taken into account when testing the Yerkes-Dodson Law.

REFERENCE

BROADHURST, P. L. Emotionality and the Yerkes-Dodson Law. *J. exp. Psychol.*, 1957, **54**, 345–352.

(Received September 23, 1959)

---

[3] P. L. Broadhurst. Personal communication, September 1958.

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

SUPPLEMENTARY REPORT: FURTHER FACTORS AFFECTING THE PROBABILITY OF CHANGING RESPONSES[1]

MAYNARD W. SHELLY II

Office of Naval Research

In a previous study (Shelly, 1958) the effects of response contingent probabilities upon response change was studied. Two factors were identified: changes are conditional upon (*a*) repetition not being reinforced, and (*b*) the probability that a change results in a reinforcement. This supplementary report extends these results.

Method.—All procedures, including numbers of *S*s per condition, were identical to those used in Shelly (1958). In the preceding study the conditions were presented by matrices, each row of which differred from the one above it through moving the elements one place to the right. Thus, to specify a condition it becomes necessary only to give the first row with the understanding that the first position specifies the probability that repetition will be reinforced. Using this method of presentation, five nonexclusive sets of conditions can be described: Set A: {(.00, 1.0, .00, .00), (.00, 1.0, 1.0, .00), (.00, 1.0, 1.0, 1.0)}; Set B: {(.00, .33, .33, .33), (.00, .50, .50, .00), (.00, .50, .50)}; Set C: {(.50, .00, .00, .00), (.50, .25, .00, .00), (.50, .50, .00, .00), (.25, .25, .25, .25), (.50, .50, .50, .50)}; Set D: {(.50, .50, .00, .00), (.50, .25, .25, .00), (.50, .16, .16, .16)}; and Set E: {S = (.75, .25, .00, .00), T = (.25, .75, .00, .00), Z = (.00, .75, .25, .00)}. For Set E, the labeling of conditions is done upon the basis of the first letter of the percentage of reinforcement associated with repetition, e.g., S for 75%. To refer to any other condition, e.g., the first condition of Set A, the notation A_1 will be used.

Results and conclusions.—Within the sets of conditions, methods of analysis identical to those used in the previous paper, including

---

[1] This research was completed at the University of Wisconsin while the author was a Social Science Research Council Research Training Fellow. Some of these results are included in the author's PhD dissertation. The author wishes to thank the Office of Naval Research for providing him with time, and the Department of Psychology, Northwestern University for the facilities to complete this research.

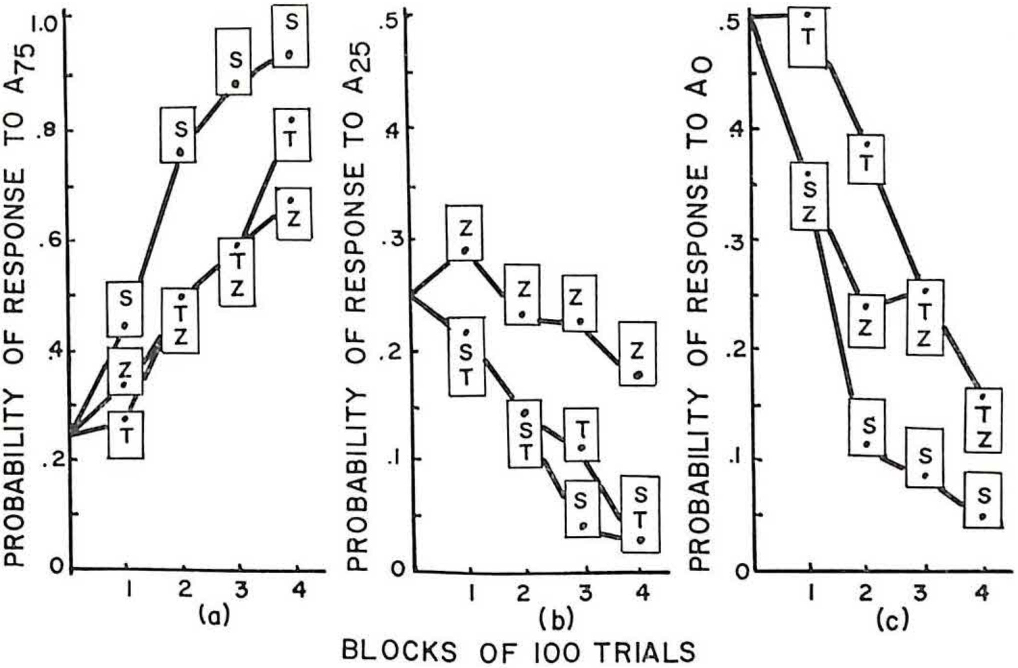


FIG. 1. Probability of responding to an alternative when repetition is reinforced with varying probabilities.

a constant confidence level of .05, will be used. "P_C" stands for the probability of changing responses on successive trials. To indicate that, for example, P_C for A_1 and A_2 was greater than P_C for A_2 on Blocks 3 and 4 of 100 trials, the short hand notation $(P_C(A_1, A_2) > P_C(A_2) | 3, 4)$ will be used.

The following are the conclusions together with the supporting results.

1. Increasing the reinforcement of repetition, even though changing responses leads to equal reinforcement, reduced P_C: $(P_C(C_4) > P_C(C_5) | 1, 2, 3, 4)$.

2. A larger reinforcement for a specific change appears to be superior in increasing P_C than the same probability spread over a number of alternatives: $(P_C(D_1) > P_C(D_3) | 2)$.

3. The differences in (2) were small, however, and when the number of alternatives to which a change can be made is increased *without* any decrease in the probability that a change can lead to a reinforcement, then P_C may increase: $(P_C(A_3) > P_C(A_1) | 1)$.

4. In agreement with Conclusion 2, when the probability of a single change leading to a win is increased *without* altering the probability of reinforcing repetition, P_C, if it changes significantly, increases: $(P_C(C_3) > P_C(C_2, C_1) | 2)$; $(P_C(C_3) > P_C(C_1) | 4)$.

5. There was some evidence that number of available alternatives had an initial effect on P_C. Although the only significant relation was $(P_C(B_1) > P_C(B_3) | 1)$, $P_C(B_2)$ lay regularly between the other two corresponding probabilities.

6. That the choice of an alternative on successive trials can be influenced by the value of the conditional probability of reinforcing repetition can be seen in Fig. 1. In this figure A_{75} refers to that alternative which on a given trial has the probability of reinforcement of .75 associated with it; corresponding interpretations apply to A_{25} and A_0. Where two values do not differ significantly for the same block of 100 trials, they are combined into a single point. If the probability of reinforcement were independent of changing or repeating responses, all three graphs would exhibit but a single curve.

The preceding results extend the findings of Shelly (1958) and would seem to indicate that factors influencing changing responses need to be considered in predicting rate of learning.

REFERENCE

SHELLY, M. W. The effects of response contingent probabilities which favor response change. *J. exp. Psychol.*, 1958, 56, 239–245.

(Received February 10, 1960)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

SUPPLEMENTARY REPORT: FAMILIARITY AND FREQUENCY[1]

CLYDE E. NOBLE

Montana State University

The writer has proposed (Noble, 1954) a hyperbolic law connecting judged familiarity (f) and frequency of occurrence (n) for unfamiliar and meaningless verbal stimuli. The empirical data, reproduced in the upper half of Fig. 1, were fitted by the equation,

$$f = \frac{n}{an + b} \qquad [1]$$

where f denotes successive-intervals transformations in z units of familiarity ratings (Noble, 1953), n is the physical number of exposures, and a and b are empirical constants. An index of correlation (p), calculated from a linear plot of reciprocals ($1/f \alpha 1/n$) by the product-moment method, gave a value based on 288 *S*s of .998 ($N = 9$). Equation 1 thus accounts for 99.6% ($p^2 \times 100$) of the variance in the dependent variable.

Repeating this experiment with random visual shapes on 600 *S*s from the same population, but using a graphical scaling method, Arnoult (1956) obtained a similar curve. To describe it, however, he chose the logarithmic function,

$$f = a \log (n + b) - c \qquad [2]$$

Arnoult reported a higher index of correlation for a semilog reduction of Equation 2 ($p = .997$), in which $b = 1$, than for the linear plot resulting from taking reciprocals of Equation 1 ($p = .950$).[2] Comparing p values from reductions of Equations 1 and 2, however, is dubious practice because mathematically different transformations are involved. One should not forget that p^2 then describes the precision of the general law modified by scale characteristics.

Goodness of fit depends, furthermore, upon the particular scaling operations followed in calculating familiarity scores and upon whether any *ad hoc* constants are added to the independent variable n. Limiting our attention to the most comparable aspects of these experiments, we find that if the same (Thurstone-Edwards) methodology used by Noble (1953, 1954) is applied to Arnoult's original frequency distributions, the rectangular hyperbola shown in the lower half of Fig. 1 results. Its p value derived from a rectification of Equation 1 is .997 ($N = 10$), accounting for 99.4% of the variance. If we apply Equation 2 in semilog form to the same data, the p value is .9917 ($N = 10$), explaining only 98.3% of the variance in familiarity (f). A still smaller index of determination might result if Arnoult's 5-hr. delay group were eliminated since their ratings tended to be higher.

When procedures for measuring the dependent variable are thus equated, Arnoult's data confirm Noble's finding that familiarity (f) is determined almost uniquely by frequency of exposure (n), and that this dependency grows according to a hyperbolic law. The precision and generality of Equation 1 is striking; only the slope (b) and intercept (a) parameters of the reduction line, $1/f = b(1/n) + a$, change when the task and stimulus characteristics are experimentally altered. Errors of measurement are reduced to the order of 1% in both cases.

Because of the "extraordinary accuracy" of the empirical fits of Equations 1 and 2, a casual reader may be tempted to conclude that these two sets of observations are "genuinely different" (Arnoult, 1956, p. 275). After descriptively asserting that "f is a monotonic, negatively accelerated function of n" (p. 272), Arnoult proceeded to embrace the notion of an infinitely-rising curve (p. 275) which is conventional among perception researchers. Hyperbolas have asymptotes, but logarithmic curves are unbounded. On the hyperbolic hypothesis, the calculated asymptote ($1/b$) for Noble's data has a maximum f value of $1/.130 = 7.69$ z units, while that for Arnoult's data has an f value of $1/.358 = 2.79$ z units. A relatively simple experimental resolution of the issue would be possible by extending the range of n values from 0 out to frequencies of 50 or 100. If Equation 1 is the correct form of the familiarity-frequency law, very little further rise should occur in the two functions as n in-

---

[1] This work was supported in part by a contract between the office of Naval Research and Montana State University. R. W. Chambers and Q. A. Leland checked the statistical analyses.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Arnoult kindly made available some of his data for reanalysis in our laboratory. His index of .950 (Arnoult, p. 274) for a linearized Equation 1 is erroneous; the value of p applied to data derived by the Garner-Hake method is actually .974 ($N = 10$).

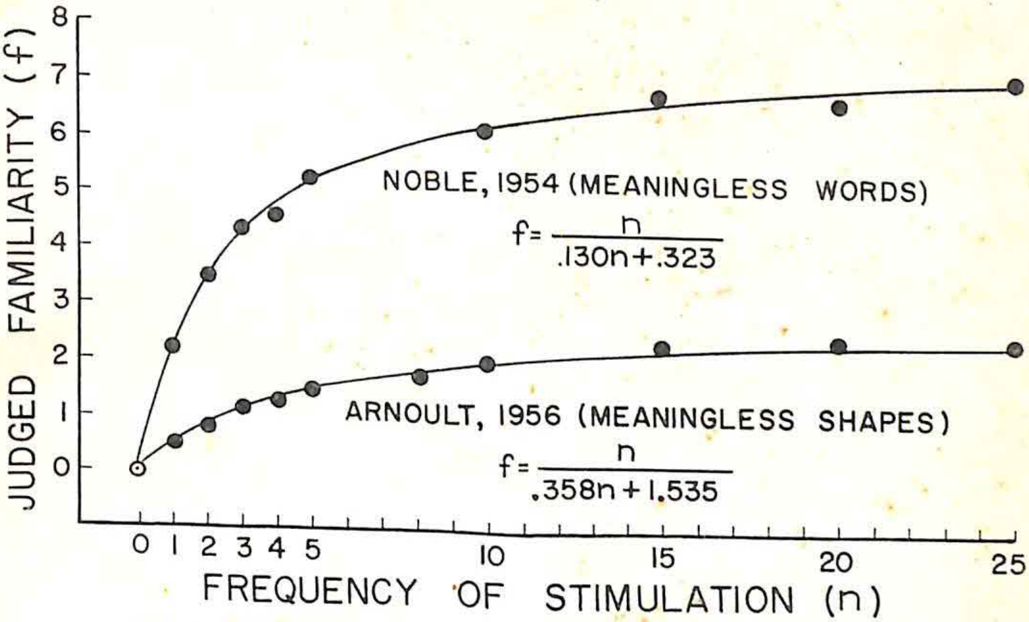


FIG. 1. The familiarity-frequency relationship for two different classes of stimuli.

creases without bound. Since the *f* scores lie on an interval scale, the 0–25 segment shown in Fig. 1 would have to be re-explored and the same *S*s should be employed at all points on the curve.

Until the range of observations is increased well beyond $n = 25$, one may continue to entertain an hypothesis (Equation 1) which predicts behavior with less than 1% error, which specifies a rational origin and finite asymptotes, which is more parsimonious than Arnoult's special logarithmic function (Equation 2), and which applies equally well to two classes of stimuli. Referring again to Fig. 1, the writer reaffirms his contention (Noble, 1953) that familiarity is a learnable attribute of stimuli, and that Fig. 1 contains acquisition curves rather than psychophysical functions. This associational hypothesis derives not from hyperbolic-logarithmic distinctions but from a consideration of the fundamental operations involved in the process of familiarization (Noble, 1954, p. 15). The S-R law represented by Equation 1 appears to hold as well for the stimulus class of random shapes (Arnoult, 1956) as for that of meaningless words (Noble, 1954). This is not true of Equation 2, whose goodness-of-fit index is only 94.65% ($N = 9$). In every instance the hyperbolic law predicts the data better than the logarithmic law when familiarity is defined by the original scaling operations (Noble, 1953).

Finally, experimental attempts to delimit the law's invariance by systematically exploring other sensory modalities should be undertaken in conjunction with theoretical studies designed to clarify the basic learning process by means of which frequency exerts its effect on familiarity.

REFERENCES

ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, **51**, 269–276.

NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.

NOBLE, C. E. The familiarity-frequency relationship. *J. exp. Psychol.*, 1954, **47**, 13–16.

(Received August 3, 1959)

Journal of Experimental Psychology
Vol. 59, No. 6, 1960

SUPPLEMENTARY REPORT: THE EFFECTS OF VERBAL REINFORCEMENT COMBINATIONS ON LEARNING IN CHILDREN[1]

CAROLYN CURRY

Indiana University

Buss and Buss (1956, Exp. I) found that using the verbal reinforcement combinations Right-Wrong (RW) and Nothing-Wrong (NW) led to faster learning than did Right-Nothing (RN), and that the former two combinations were approximately equally effective. The present experiment is an extension of the previous study using a new population (normal children) and a new response measure (errors in a fixed series of trials).

Method.—The same stimulus materials and procedure were used as in the previous study with the following exceptions. Each *S* was given 150 cards to sort with correct sorting for the first 50 cards on the basis of color and for the next 100 cards on the basis of number. Ninety grade school pupils were assigned to three groups of boys and three groups of girls. Each group was matched for IQ level on the basis of the Otis Quick Scoring Test of Mental Ability, and each group contained identical proportions of Grades 5 and 6. One group of each sex was assigned to each of the three verbal reinforcement combinations.

Results.—The means and *SD*s of the total number of errors made by the three treatment groups during each sorting are shown in Table 1. These data were subjected to analyses of variance (Treatments × IQ level × Sex) with each basis for sorting being considered separately. The effects of Treatments were significant at the .01 level for color-sorting and at the .05 level for number-sorting. A sex difference was also significant at the .05 level for color-sorting, the boys making fewer errors than the girls. No other differences were significant. For both color- and number-sorting, *t* tests revealed that the RN group was significantly different from the NW and RW groups. The latter two groups did not differ significantly in either instance.

TABLE 1

MEANS AND *SD*s OF ERRORS FOR VERBAL REINFORCEMENT COMBINATION GROUPS

Group	Color-sorting		Number-sorting	
	Mean	*SD*	Mean	*SD*
RW	14.83	17.45	28.83	32.97
RN	26.87	23.00	44.90	43.93
NW	14.16	17.16	27.97	32.68

Discussion.—The results agree with those obtained by Buss and Buss (1956) for adults on a similar task. However, the results are at variance with those reported by Terrell and Kennedy (1957). These investigators studied the effects of a number of types of reinforcement on learning in children. Among these were *E*'s verbalizations to indicate only correct responses for one group and only incorrect responses for a second group. Although these two groups are somewhat comparable to the RN and NW groups of the present experiment, there were no significant differences in the effectiveness of these two combinations. However, the Terrell and Kennedy procedure also included a light flash which indicated correct responses to all *S*s. In the present experiment, and in that reported by Buss and Buss, the *S*'s only source of information about the correctness of a response was the assigned verbal reinforcement combination. The difference between these results and those obtained by Terrell and Kennedy may be a result of the presence of additional information provided to *S*s about their responses in the latter experiment. It seems possible that the differential effectiveness of the verbal reinforcement combinations may be dependent on there being no other information about the correctness of a response available.

REFERENCES

BUSS, A. H., & BUSS, E. H. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, **52**, 283–287.

TERRELL, G., JR., & KENNEDY, W. A. The discrimination learning and transposition in children as a function of the nature of the reward. *J. exp. Psychol.*, 1957, **53**, 257–260.

(Received November 19, 1959)

---

[1] This paper is based on an M.A. thesis submitted to the graduate school of Indiana University. The author wishes to express her appreciation to A. M. Buchwald, thesis chairman, for his many helpful suggestions. The major results of this study were presented at the 1959 meeting of the APA.

AU OF GOVT. & PSY [illegible]
LIBRARY
[illegible]